EDUCATION PRODUCTION FUNCTIONS IN POLICY MAKING: A CRITICAL ANALYSIS

by

Aubrey Hampton Price

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Educational Administration

APPROVED:

Richard Salmon, Chairman

Glen Earthman

Kern Alexander

Jimmie Fortune

James Rountree

May, 1994

Blacksburg, Virginia

EDUCATION PRODUCTION FUNCTIONS IN POLICY MAKING: A CRITICAL ANALYSIS

by

Aubrey Hampton Price

Committee Chairman: Richard G. Salmon
Educational Administration

(ABSTRACT)

The conceptual, methodological, and empirical issues of education production function analysis were examined. Specific issues relating to model design, proxy and variable selection, levels of analysis, standards of adequacy, and implications for policy were addressed.

Selected studies and dissertations published since the Equal Education Opportunity Report that identified any positive significant relationship between financial resources and student achievement were evaluated. Each study was examined with reference to conceptual and empirical problems of analysis. Specifically, the researcher identified measures of dependent and independent variables, the level of data aggregation, the direction and effect of independent variables, characteristics of the sample, the statistical methodology employed in the analysis, and the conclusions of the study.

This research discussed factors that result in discrepant findings in the relationship between expenditures and achievement. From this perspective, conclusions about the use of education production function studies in assessing this

relationship were stated. The examination of empirical studies provided a context for the identification of conceptual, methodological, and data issues necessary to assess the potential of production function analyses to contribute to policy decisions.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my graduate committee for their patience and support during the dissertation process. I especially thank Richard Salmon for his guidance and encouragement while chairing the dissertation and Glen Earthman for assisting me in many ways over the years. Jimmie Fortune and Kern Alexander were instrumental in helping me select this project. James Rountree has offered valuable suggestions on this enterprise and other facets of the program. I would like to express gratitude to David Parks and Wayne Worner for the role each has played in my education.

I have been fortunate to work with many fine teachers and staff personnel during my graduate program. To my friends from Mary Hunter, Fieldale, Mt. Olivet, and Virginia Tech, I extend my sincerest appreciation. Finally, I must express my deepest gratitude to my wife, Judy, and sons, David and Brian, for sharing this long journey to its completion.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Production-function studies in education have a thirty year history. The production function or input-output analysis is a model of relationships based on the economic theory of the firm. It requires the measurement of precise increments of resource inputs that yield specific increments of production outputs. The model is a valuable tool for planning industrial production. The unity of theoretical design, empirical analysis, and practical result has been basic to the development of the production function model. In education, however, wide variation in the use of the model in empirical studies under very diverse conditions has resulted in confusing and contradictory results. Unfortunately, premature claims about the efficacy of this analytical tool for policy making can result in counterproductive decisions that reduce the efficiency and productiveness of schools (Monk, 1990; Berliner, 1993). Yet, policy makers cannot ignore the opportunity to develop a more precise understanding of the links between resource investment and achievement in education.

Education production function analysis began with the Equal Educational Opportunity Report (1966), cited as The Coleman Report for its principal author. Since the project was funded by Section 402 of the Civil Rights Act of 1964, federal officials attempted to make the data base one of the most comprehensive collections of information about public schools available. With these data they addressed comprehensive questions about the effect of schools and other factors on student achievement. The four major questions examined by the study were:

1. The extent to which racial groups were segregated in the public schools.
2. The presence of equal educational opportunities by criteria regarded as indicators of quality: tangible characteristics of schools in physical facilities; curriculum areas such as high school tracking; characteristics of teachers including salary, verbal ability, experience, and attitudes; student body characteristics including self-attitudes and academic goals, socioeconomic status (SES), and parent education levels.
3. How much students learn as measured by standardized tests.
4. The relationship between what students learn and the kinds of schools they attend (Coleman, et al., 1966).

These objectives covered most of the areas investigated by the education production function analysis. Since many readers interpreted the study to indicate that schools had little influence on student achievement independent of family background and general social context, it provided great impetus for research (see, for example, Mostellar & Moynihan, 1972). This report, and the debate it began, have been the catalyst for much of the research into education cost-quality issues which have employed production function analyses (Coons, Clune, & Sugarman, 1970; Hanushek, 1991).

Questions about the influence of financial resources on the achievement of students that emerged from the Coleman Report lie at the heart of the two dramas now playing on the stage of education: one is the national reform effort that began with A Nation at Risk; the other is the wave of state litigation that seeks to reshape education finance and distribute resources more equally to all students (The State of Inequality, 1991). In both dramas education production function analysis has provided information to justify the perspectives and decisions of policy makers. For example, during his tenure as secretary of education, William Bennett often referred to research that proved there was no relationship between school expenditures and student achievement. According to Baker (1991), Bennett was referring to Eric Hanushek's 1986 education production function chart. Also, recent equity suits have involved the litigants in arguments about the importance of financial resources based on education production function analyses (Testimony, 1987; Testimony, 1991; Fortune, 1992). These examples show the important role such studies have played, and continue to play, in federal and state education finance policies.

JUSTIFICATION FOR THE STUDY

Production function analysis has a proven validity for industrial production that lends credibility to its transfer into other disciplines. The use of aggregate data in statistical analysis to predict performance or identify correlations, trends, or causality in basic scientific research has encouraged social scientists to emulate their peers. When administrative agencies, courts, and legislatures turn to the experts for informed decisions about spending money on education, the education production function study may appear to provide the best evidence available from the research community since education researchers have published numerous studies using this methodology. The opportunity for researchers to influence policy, however, can overwhelm the nuances of statistical analysis and the interpretive limitations imposed by nonexperimental research.

Therefore, it is desirable, even necessary, that the education production function model be examined critically.

PURPOSE OF THE STUDY

Previous studies of education production functions have focused on the statistical influence of specific independent variables, or correlated the significance of several variables over many studies. The variation in significance and direction of effect among variables aggregated in this manner has tended to look beyond the framework and setting of each study. Because the instructional environment is highly idiosyncratic, this approach may ignore patterns that exist in the context of the samples. This analysis will examine studies that have positive and significant correlations or regression coefficients to identify the patterns that exist between instructional expenditures and student achievement. In addition, this study will identify conceptual, methodological, and empirical issues of education production function analysis. With this information the selected studies will be examined in the context of the generic problems of analysis. The findings from this evaluation will provide evidence to discuss the conditions under which education production function studies show that instructional expenditures correlate with student achievement. Examination of these issues will permit the discussion of conceptualization and methodological and data needs which could allow the production function analysis to assist policy makers in the distribution of funds.

STATEMENT OF THE ANALYTICAL APPROACH

The following criteria will be used to analyze the selected studies. The researcher will:

1. Identify and explain components of the conceptual model.
2. Identify and discuss statistical issues and problems related to empirical studies.
3. Review empirical studies at the K-12 level that incorporate expenditure-related items as an independent variables by utilizing the following criteria:
   A. Identify specific measures of dependent and independent variables.
   B. Identify the level of aggregation for variables.
   C. Identify the direction and effect of independent variables.
   D. Identify the sample size and source for variables.
   E. Identify the statistical methodology employed in the analysis.
4. Identify conditions under which financial resources appear to influence student achievement.
5. Draw conclusions about the use of education production function studies for educational decision making.

SELECTION OF STUDIES

Studies will be selected from journals, dissertations, and ERIC that focus on the K-12 education environment conducted since the Coleman Report. Only those that address the direct or indirect effect of expenditures on student achievement and report one or more positive and significant correlations or regression coefficients will be examined. To qualify for inclusion as an education production function analysis, each study must have at least one input and one output variable, correlation data for the input and output variables, the assumption of a link between school expenditures and student achievement at the given level of data aggregation, and the assumption of unbiased estimates for variables.

ORGANIZATION OF THE STUDY

This research is organized into six chapters. Chapter I contains the introduction, justification of the study, purpose of the study, statement of the analytical approach, criteria for the selection of studies, and the organization of the study.

Chapter II consists of a discussion of the conceptual model and comparisons between the industrial model and the education model.

Chapter III provides a discussion of the issues surrounding empirical studies from a review of the literature.

Chapter IV presents a description of the selected studies and the data provided from the analyses.

Chapter V presents the findings of the study that identify the patterns of relationships between school expenditures and achievement.

Chapter VI draws conclusions from the findings, states implications for production function analysis, and makes recommendations for future studies.

CHAPTER II

THE CONCEPTUAL MODEL

The Industrial Model

"The goal of the entrepreneur is to maximize the quantity of his output for a given cost level, but generally his cost is variable, and he desires to maximize his profit" (Henderson & Quandt, p. 53). The production function or input-output analysis is a model of the economic relationship between the maximum amount of output that can be produced and the inputs required to make that output. The model is defined for given technological levels that are subject to change differentially. When linked with cost analyses, the production model can be used to measure increments of resource inputs that yield specific increments of production outputs at precise cost estimates. The model is a valuable tool for efficiently planning industrial production based on least cost input estimates. "There are thousands of different production functions in the American economy: at least one for each firm and product" (Samuelson & Nordhaus, 1989, p. 499).

In the early development of the production model, Cobb and Douglas (1928) tried to measure changes in the amount of labor and capital used to create a given amount of goods and to determine the relationships between the product, labor, and capital. They quantified the components for a given time and estimated the relationship between factors using regression analysis. Although data collection in the late nineteenth and early twentieth centuries was imperfect, Cobb and Douglas compared the results of their analytical model with information provided from federal government sources enabling the concept to be validated. Today, of course, cost and quantity data are abundant for measuring the physical factors in production.

The Education Model

The goal of the education model is to explain the relationships between inputs and outputs that would permit the maximization of student achievement from a given combination of input factors. This approach requires the use of regression analysis for explanatory purposes initially. Such a use of regression requires the specification of a conceptual model from which the data can be meaningfully interpreted (Pedhazur, 1982). Variables in the education equation may include some measures of the individual's ability, family, peer, financial, and school inputs with a measure of student achievement as the output factor.

Hanushek (1972) developed the following linear model in which achievement (A) cumulative to time (t) is equal to a vector of family characteristics cumulative to time (F^t), a vector of peer group influences over time (P^t), a vector of initial genetic endowments (I), and a vector of school inputs cumulative to time (S^t):

$$A_t = g(F^t, P^t, I, S^t).$$

Since precise measures for the inputs in each vector do not exist, proxies developed and collected for other purposes must be entered into the regression equation. Proxies are measurable variables that are assumed to correlate highly with a latent, unmeasurable variable identified in the model. Unfortunately, by definition, true variables cannot be identified from proxies, only estimated. According to Hanushek, the qualities and interactions considered in the true variables are not normally purchased in the marketplace. The marginal products identified in the conceptual model refer to the true variables--those which yield output changes when altered, not to proxy variables. Therefore, a true education production function is not and, under present circumstances, cannot be known. Any policy decision based on an estimated production function is clearly speculative.

The proxies for each variable (vector) in the model represent a broad range of demographic and economic data.

1. Family background: "income level, SES, parents' education, number of people in the family, the family structure, languages spoken in the home, size of home, indices of parent interest in education, and so forth" (p. 27).
2. Peer influence: "appropriately aggregated vectors of the attitudes, backgrounds, and performances of other students and children which the individual comes into contact with. Empirically this implies using aggregates of the proxy measures of home environment" (p. 27).
3. Initial endowments: ". . . in empirical work, no measure of initial endowments is available" (p. 30).
4. School factors: "teacher characteristics and attitudes, physical characteristics of the school, curriculum, etc. . . . school inputs to the individual should be analyzed" (p. 31).

5. Achievement measures: mean verbal and mathematics scores on standardized tests were used for separate analyses. The California single system sample used the Stanford Achievement test. The Coleman researchers relied upon the verbal score from SCAT and the mathematics score from STEP.

In empirical analysis the model is significantly altered. Hanushek's analysis from the 1972 study (using Coleman data) incorporated no teacher experience variable. There is also a fundamental problem with the use of socioeconomic (SES) data to subsume the conceptual element of initial endowments (intelligence). If cognitive ability is heterogeneously distributed by nature as assumed in the normal curve, the model is confounded by reliance solely on SES measures. Since there is no model variable for family background in this analysis, the conceptual design is essentially dismissed for lack of data. These problems exist even in the circumstance where "data were specified and collected for the sole purpose of looking at the educational process instead of attempting to squeeze data gathered for different purposes into the framework of the conceptual model" (p. 36).

The way around these dilemmas is the "value added" approach (Bowles, 1970; Hanushek, 1972; Harbison & Hanushek, 1992). While Hanushek (1972) uses the procedure in his single school, individual student analysis, he does not extol its virtues so fully as he does in later publications (See, for example, <u>Education Performance of the Poor</u>, 1992). As a practical matter for Hanushek, it becomes the only available methodology that can allow for the incorporation of missing variables from the conceptual model in the empirical application. This solution creates its own dilemmas that are addressed at the end of this section.

In Education Performance of the Poor (1992), Harbison and Hanushek presented a revised conceptual model:

$$A_t = f(\ F_{(t)},\ S_{(t)},\ O_{(t)},\ e_t\ )$$

where F(t) = a vector of the student's family background and family educational inputs cumulative to time;

S(t) = a vector of the student's teacher and school inputs cumulative to time, t;

O(t) = a vector of other relevant inputs such as community factors and peers cumulative to time, t;

e_t = unmeasured factors that add to achievement at time t.

The model incorporates an error term, e_t, to indicate that we can never observe all the factors affecting achievement. The error term has a random and unpredictable effect on the estimation and interpretation of other factors in the model.

According to Hanushek, when the vector of various school factors denoted by $S(^t)$ includes the pertinent instruments of policy, the relative effectiveness of possible educational strategies can be compared both with each other and potentially with the costs.

Under experimental conditions it would be possible to estimate overall mean achievement differences between schools after accounting for measurable resource and family differences. Second, it would be possible to measure explicitly the specific school resources provided by the project and to include these factors in estimates of student achievement relationships. This would provide regression coefficients that can be used in assessing changes in performance produced by the specific inputs of the project.

The most serious issue is the probability of getting biased statistical estimates of the effectiveness of different school resources. In the equation, the error term, e_t, includes all unmeasured influences on achievement and represents the source of bias. It is natural to expect many influences on individual achievement to be unmeasured. The key issue is whether the collection of these factors is unrelated to the influences on achievement that are measured and included in the analysis. If unrelated, standard regression analysis provides unbiased estimates of the achievement relationships. If they are systematically related, however, the parameter estimates will be biased and their use for evaluation or policy analysis will tend to be misleading. It is difficult to accept that these error terms are uncorrelated with the measured inputs to achievement. These error terms are likely to contain a variety of unmeasured factors that are systematic. First, since education is a cumulative process, the entire history of inputs is necessary to characterize achievement at any point in time. Second, most research designs limit the range and character of the observed data. Even with very thorough research designs, it is difficult to record any qualitative differences in teacher behaviors. Even where qualitative observations have been tried, student responses have been almost impossible to chart. Third, some factors nearly defy measurement. For example, differences in the intelligence of students are important in determining achievement differences. Yet, there is little consensus on how intelligence might be measured, and available instruments are not easy to administer to large numbers of students. The use of "verbal ability" as a measure of intelligence in some surveys is confusing. (The National Opinion Research Center, Annual Survey, includes such a measure that is sometimes used as a proxy for intelligence.) The Coleman Report used "verbal ability" as a measure of achievement. In addition, the motivation and aspirations of children are extraordinarily difficult to measure even though they are important (Berliner, 1993).

These unmeasured factors are likely to be correlated with measured family and school variables. Past school experiences tend to be related to family characteristics and school inputs; qualitative differences in inputs often correspond to quantity and family choices; the abilities, motivations, and aspirations of students tend to be correlated with measured family characteristics. "The risk of biased parameter estimates--and unreliable policy conclusions--is thus substantial" (Harbison & Hanushek, 1992, pp. 85-86).

According to Hanushek, it is less problematic to focus on gain scores in what he labels a "value added approach," A_t-A_t*. Gain scores require repeated sampling of the same individuals. (This issue is explored in another context in the section on the influence of time.) The analysis of gain scores, or "the value added approach," has a considerable body of literature (Bloom, 1964; Campbell, 1963). Such gains are subject to a variety of internal validity challenges. Regression to the mean, history, maturation, testing effects, and mortality are all problems associated with gain scores where neither randomization nor control groups are used in research--the standard research conditions in public school settings (Huck, Cormier, & Bounds, 1974; Harris, 1963). Even more significantly, the assumption of linearity becomes highly problematic where human performance is associated with time factors (Wolfle, 1993A; Bloom, 1964). Curvilinear models become more the norm as a function of Samuelson's "empirical regularities." (See the discussion of the economic model.) Therefore, Hanushek's conceptual linear model diminishes in validity in a "value added" context and a quadratic model becomes more explanatory of learning characteristics over time.

Comparison of the Models

Several features of the economic model deserve attention because of differences with the education design. The marginal product of an input is the additional output produced by a unit of that input when all other inputs are held constant. It is a function of the law of diminishing returns. "The law of diminishing returns holds that the marginal product of each unit of input will decline as the amount of that input increases, holding all other inputs constant" (Samuelson & Nordhaus, 1989, p. 501). The dependency on teachers (hired inputs) epitomizes a reliance on one factor of production that economists would find alarming (Monk, 1990). Without alternative instructional inputs being available, the theoretical probability is that the marginal product will decline. The education model, however, assumes a linear relationship between inputs and output(s) in which there is a constant and unchanging marginal product. In economic analysis, the marginal product displays a curvilinear relationship to the input. The scarcity of curvilinear relationships in education studies has been addressed elsewhere in this analysis. It is important to identify these fundamental deviations between the industrial and the education models so that policy makers do not assume a false equality between them (See Bowles, 1970; Hanushek, 1979; Bridge, Judd, & Moock, 1979).

Returns to scale refers to the "responsiveness of total product when *all* the inputs are increased *proportionately*" (Samuelson & Nordhaus, 1989, p. 503). Economists usually seek to identify increasing, decreasing, or constant returns to scale. Constant returns to scale is assumed to be attainable in most production activities (Samuelson & Nordhaus, 1989).

In education attempts to increase returns to scale have been addressed through consolidation of smaller districts into larger ones (Verstegen, 1990; Walberg, 1987). The research suggests that larger districts produce lower achievement than smaller districts (Verstegen, 1990; Walberg, 1987; Butler & Monk, 1985). The per pupil expenditures in large cities, however, must address dysfunctional socioeconomic conditions that smaller localities may not have. This comparison of larger and smaller school districts, therefore, may often ignore conditions within the districts that work against achievement. For example, the Fairfax County system in Virginia (127,000 ADM in 1989-90) is the largest in the state and produces an enviable record of achievement (Outcome accountability project, 1992). Socioeconomic factors in this district strongly favor high investment in education and equally high student achievement. Throughout the country, however, per pupil costs in the range of district sizes from 500 to 5000 tend to differ very little (Turner, Camilli, Kroc, & Hoover, 1986; Walberg, 1992). These results suggests that constant or increasing returns to scale is not the norm in education and illustrate another area in which education and standard production functions diverge in substantive ways.

The substitutability of elements in the production function model is a rationale for its use in manufacturing and industry. If the price of one factor falls while other factor prices remain the same, substituting the lower-priced input for another could allow for a reduction in costs without any loss of quantity or quality (Samuelson & Nordhaus, 1989). To do this, it is necessary to identify inputs that contribute to outcomes; determine how much difference each input makes and their costs; and substitute effective resources for less effective ones (Bridge, Judd, & Moock, 1979). The opportunity to evaluate the substitution of elements in the education model is severely limited. Children require human supervision and guidance restricting the possibilities of the substitution of capital for labor. Beyond this condition, however, it is a consequence of the limitations imposed on experimental research within the school setting. From another perspective, it represents the inflexibility of the education model. If classroom level analysis were developed to the point that teacher effects could be assessed, for example, then alternative classroom level inputs might be evaluable. Finally, problems with substitutability among education inputs is a consequence of the interpretation of analytical results. "If we find that two variables--an input and an outcome--are unrelated statistically it may be because (a) the data are accurate and in fact the two variables are unrelated; (b) the variables are related but one (or both) of the measures is insensitive (not detecting variance that actually exists) or is producing random values; or (c) the variables are unrelated as the data show, but we have failed to measure a relevant outcome that the input does contribute to in some significant manner" (Bridge, Judd, & Moock, 1979, p. 26).

Teachers remain the most important schooling input. The addition of electronic teaching aids, such as computers, has not reduced the need for professional instruction. The use of paraprofessionals to replace teachers is as close to the concept of substitutability as education has reached. The substitution of longer hours in the school day or the extension of the school year into the summer to more efficiently use facilities represent efforts to substitute resources--traditional down time for building use substituted for the cost of additional facility construction. Unfortunately, the inability to provide information on resource arrangements within the education model limits the value of the production function study to meet the needs of decision makers.

Technological change is the "invention of new products, improvements in old products, or changes in the processes for producing goods and services" (Samuelson & Nordhaus, 1989, p. 506). In education, technology encompasses the methodology of instruction (processes) as well as the hardware of computers and other media. Education has remained labor intensive as other industries have replaced people with machines. The inability to substitute capital for labor in education makes teacher training a primary means to transform the process. Teacher behaviors, however, lie outside the ability of the production function analysis to evaluate. Therefore, crucial policy information that could transform the education system, also, lies outside the purview of production function study (See Walberg, 1992, for instructional strategies that demonstrate significant effects on student achievement).

Since the education model lacks direct measures of the variables identified in the conceptual design and proxies must be selected from available data to serve as substitutes in a regression equation, there is no opportunity to validate the model against data supplied from other sources. A comparison of the estimated production model cannot be made with the true inputs since only proxy data are available. Using this kind of comparison with true data enabled Cobb and Douglas to validate and refine the industrial model (See Cobb & Douglas, 1928, p. 153). The education model, however it may be defined, remains unverified by any objective methodology.

Costing variables is a relatively simple task for calculations using the industrial model. Data are collected specifically for that purpose by government agencies and industries. The Cobb-Douglas equation focuses on physical inputs and outputs primarily. The costing of variables for the education model remains in a primitive condition, partially because inputs and outputs are neither exclusively nor most importantly physical. Efforts such as that of Rossmiller (1986) and Ferguson (1991) remain relatively unmatched examples of the possibilities of costing variables for American schools. The variability of the results of these studies are sufficient to give policy makers pause at the cost effectiveness of such production function studies. (The Vanderbilt computer simulation, ITCOT, is an innovative approach to educating administrators by linking reform scenarios to time and financial costs.)

Human variables play a significantly different role in the industrial and education models. Cobb and Douglas recognized that neither the quality of labor nor the intensity of work could be measured quantitatively in their initial application of the model. This is a reasonable certainty for the education model as well. Nevertheless, labor inputs could be quantified and held as a fixed or variable factor to determine relative input value since other measures were easily quantifiable. Teacher and student inputs remain less easily quantified. The student is both an input and an output factor. In the industrial model the output commodity is not an active participant on the input side. Even economists acknowledge that the acquisitive behavior ascribed to mankind by the assumptions of the discipline often fails as an explanatory tool (Weisskopf, 1973). The dual role of the student in the model adds a confounding characteristic to education production. If one could assume with the economists that student acquisitiveness was maximized (for knowledge, in the education model, as the equivalent to wealth in the economic model), the problem of student behavior would be less of a conundrum. Such an assumption is difficult to sustain in the K-12 environment where production function studies have been concentrated.

Time As A Cumulative Variable

The use of time in the economic model is recognized as a continuum during which certain measurable material conditions exist (Cobb & Douglas, 1928). It is not an independent factor in production. Time affects the definition of an input as fixed or variable (changeable) based on the length of the production planning period (Henderson & Quandt, 1971). In education, time is a contributor to the product of student achievement. "Time inputs are important determinants of cognitive achievement" (Lau, 1979, p. 42). Lau can describe quantitative proxies for time inputs, but the estimate of cumulative effect is unspecified for the short time period analyzed by most education studies. The relationship of initial student endowments to achievement in an academic year, for example, can be measured. This may offer a means to evade the problem of cumulative effect. Since important time factors remain in the control of students (homework time for specific subjects and television time), reliable measures are elusive. Unfortunately, measures of initial endowments are often excluded from equations. IQ may be challenged as a baseline, because it is demonstrably a changeable measure and time may have a significant unmeasured cumulative effect upon it (Berliner, 1993). As previously discussed, Bowles (1970) and Hanushek (1972) have suggested that using the "value added" approach to measuring achievement could include all previous conditions of SES, schooling, and ability in the initial measure.

CHAPTER III

ISSUES IN EMPIRICAL STUDIES

Deficiencies of Problem Assessment

Production function analyses have not generally addressed several critical problems (Bridge, Judd, & Moock, 1979). While outputs vary widely by individuals, school inputs often may not vary greatly in either quantity or quality. When there are small input differences, the effect of the input may not be statistically significant because of the attenuation of variance in the measurement of samples. An example provided by the measurement of the effects of facilities illustrates this point. Hanushek (Harbison & Hanushek, 1992) found statistically significant effects for facilities on student achievement in his Brazilian analysis. He found no such relationship in his earlier studies in the United States (Hanushek, 1972). The differences in the resources available in the two settings provide statistical evidence that money does matter in facilities investment where there is a significant variance in the measures compared. Where the variance in facilities is reduced (attenuated) by limited measurable differences between the samples, no statistical significance is found. Production function analyses have rarely addressed attenuated variance, although it is likely to be prevalent in district level analyses within states that provide moderate equalization in funding. Pass/fail measurements of student achievement are particularly vulnerable to the problem of attenuated variance (see Price, Strickland, & Fortune, 1993).

The interaction between inputs can have an effect on output measures. Bronfenbrenner (1974) demonstrated that preschool intervention involving parents and schools can have measurable effects on IQ scores, but interventions that did not involve parents failed to produce measurable effects.

According to Piaget, the sequence of inputs can be important in the acquisition of knowledge. Attempts to teach reading before the child is "ready" will fail. Accordingly, failure to provide learning experiences at the suitable time can limit a child's development. Ethologists and psychologists provide evidence that suggests "windows of opportunity," "readiness," or "critical periods" in the development of individuals (Gray, 1958; Erikson, 1963). Production function studies have been much too limited in scope to examine such issues. The results of a production function analysis can fail to assess the presence or absence of important time-related learning inputs that have significant effects on achievement.

The need for the application of a certain amount of an input before it produces measurable results is a threshold effect (Bridge, Judd, & Moock, 1979). The analogy to the "take-off" period in Walt Rostow's <u>The Stages of Economic Growth</u> (1969) is a useful construct. Certain material preconditions must develop before a society can move into a sustained growth economy. Before these conditions are met, a ceiling exists on the attainable output (p. 4). Phillips and Marble (1986) surveyed 1,548 farmers in Guatemala. The empirical results suggested that a threshold effect of four or more years of education is necessary to produce a measurable effect on agricultural production. Recent analyses by Fortune (1992) suggest that a threshold of $600 to $700 may be linked to achievement differences in otherwise homogeneous school districts in several states. While he is careful to explain that such findings are frequently lost in the analysis of very diverse systems and the amount of money is likely to vary from state to state, the evidence of a threshold effect that relates student achievement to financial resources is a new and interesting perspective. Few studies have examined these possibilities.

Many human performance variables are related to chronological age. In statistical analysis these tend to be curvilinear relationships (Wolfle, 1993A; Bloom, 1964). Since most education production function studies are based on linear assumptions, these relationships have rarely been addressed. Economists assume curvilinear relationships to be "empirical regularities" (Samuelson & Nordhaus, 1979). The Cobb-Douglas function incorporates multiplicative equations to account for these. If the fixed relationships of material objects require curvilinear analysis, can the complexity of human interactions be subsumed by a linear model? The education model will be unable to develop a mature conceptualization without standardizing this level of sophistication.

The *assumption* of linearity that has characterized much of the empirical research limits empirical analysis. A brief example from social survey research will illustrate the nonlinearity of a simple relationship in education analysis. Using data (N=602) from the NORC survey of 1991, a regression analysis was used to test the relationship between scores on a vocabulary test and the number of hours of television viewed. Controlling for other variables in the model, the quadratic equation explained more than one and one-half times the variance in vocabulary performance compared to the linear equation; the difference in percent of variance explained was statistically significant at the .01 level.

This commonsense relationship helps demonstrate the need for curvilinear analysis in assessing the impact of resources on learning. It is possible, however, that the linear relationship is satisfactory to explain some relationships. It is sufficient to state that studies failing to test quadratic equations may be challenged on that basis alone.

Variations in Variables

Most education production function analyses select from a common set of variables. Studies include some or all of the following on the input side of the model formula: characteristics of the home or family, the community, peers, individual students, and the school or district.

The Coleman Report (1966) used home background (family), teacher, student body, and facility measures to correlate with student standardized test performance (verbal ability). Hanushek (1972) identified individual and family characteristics in one set of influences, peer characteristics, the initial endowments of students (ability), and school inputs. Standardized test performance data for mathematics and reading were output measures. Ferguson (1991) does not include variables that measure innate abilities, nor does he separate family and community characteristics on the input side. On the output side he used reading and mathematics scores from the Texas minimum skills assessment. Rossmiller (1986) used community, school district, school, student, and teacher characteristics on the input side of the equation. Stanford achievement scores in mathematics and reading were the output measures. This study also used measures from the Self-Observation Scale to examine financial relationships to student self-concept. He used affective measures on the output side of the equation.

This brief review of the literature demonstrates the variation of model design, particularly on the input side. These examples should help indicate the current dilemma in selecting appropriate measures that have impact on student achievement. Variable selection contributes to the discrepancies that exist in the results of empirical studies, although it is by no means the sole explanatory element.

Variations in Proxies

Selecting variables for a production model is a decision complicated by the need to measure the variables. No one simple, direct and complete measurement exits for family characteristics or any of the other variables. Consequently, a variety of existing measures or proxies are used to operationalize the constructs. The following lists taken primarily from Lau (1979) is reasonably thorough.

Home background (family characteristics) may include: IQ, age, education, income, occupation, race, religion, attitudes, expectations, preferences, values of either or both parents, the structure and stability of the family, the number of siblings, and locational stability. Indicators of the home learning environment may include: possession of books, radios, televisions, frequency of travel, and knowledge of foreign languages. Direct home inputs to the learning process could include the time that either or both parents, an older sibling, or even a tutor devoted to the instruction of the students being assessed.

School characteristics include the type of school (public, parochial, private), type of curriculum, level (K-12), size, ethnic composition of enrollment, class size, per pupil expenditure, or method of instruction. Direct inputs may include facilities or specific features such a science laboratories or libraries; administrators--quantity, quality (degree, experience, or other accomplishments), attitudes, and time inputs to various activities; teachers--quantity, quality, experience, attitudes and other attributes (including IQ or verbal ability, knowledge of subject, ability to communicate, responsiveness to questions), and teacher time inputs to preparation, lecturing, student consultation, and grading. Instructional methodology emphasized in the school must be measured and may include: traditional forms, instructional television, computer assisted instruction, or variations and combinations of these or other technical processes.

Community characteristics are intended to measure the degree of support for education. The type of neighborhood, size of the city, degree of community interest, attitudes, average age, mean education level, mean income level, SES (described by some measure of community wealth), and property value are among the measures selected. In some studies the peer group variable is incorporated into community characteristics.

Student characteristics include: measures of ability (IQ), age, aptitudes, race, sex, birth order, number of siblings, and previous educational achievements on standardized tests. Measures of subjective conditions that may affect achievement such as attitudes, interests, motivation, self concept, self expectations, and values have been used. Student time inputs to homework, class attendance, laboratory work, or self-study have been used.

Peer group characteristics are selected from the proxies for student and community characteristics. The influence of peers is included because of the effect on individual student achievement. These characteristics also influence teacher classroom behavior. These proxies may shift from one category to another in different studies. SES is determined by another measure such as personal income or percent of free lunch depending on whether the SES of the student body (peers) or the community is the variable. Without belaboring the obvious, studies differ on the proxies selected for variables. The results of an analysis can be significantly influenced by the proxies selected for variables.

Statistical Procedures and Problems

Nearly all education production function studies use a variation of regression analysis (See Thompson & Correa, 1989; MacPhail-Wilcox & King, 1986; Glasman & Biniaminov, 1981; Bridge, Judd, & Moock, 1979). Linear model equations (ordinary least squares) have been the predominant tools of analysis. Stepwise regression, variance-partitioning, commonality analysis, and path analysis have been used. Simultaneous equations (two-stage least squares) procedures have been employed with greater frequency recently. Ferguson (1991) used this procedure in his analysis of Texas data. Thompson and Correa (1989) tested for significant differences between groups before utilizing regression procedures. Subhypotheses were tested using four-way analysis of variance.

No amount of statistical manipulation can overcome the use of insufficient or deficient data. While Pedhazur (1982) relates this statement to the problem of multicollinearity, it has application across the range of education production functions. Regression analysis is not the problem in these studies, it is the use to which the model is subjected that creates the problems. The conceptual model has been identified in several guises in previous sections. The education model, when defined operationally for an empirical analysis, may explain **at best** a very small amount of the actual variance in student performance (Bridge, Judd, & Moock, 1979).

Regression procedures postulate certain assumptions (this discussion is based on Pedhazur, 1982) about the data used in an analysis that must be met if the results are to be trusted to provide accurate information:

1. Independent variables are fixed. The same values must be used in a replication of a study as used in the initial analysis.
2. The independent variable, X, is measured without error.
3. The dependent variable, Y, is a random variable that has a range of possible values.
4. The regression of Y on X is assumed to be linear.
5. The mean of errors for each observation of the dependent variable is zero.

6. Errors associated with one observation of the dependent variable are not correlated with errors associated with any other observation of the dependent variable.

7. The errors are assumed not to be correlated with an independent variable.

8. Errors are normally distributed.

"It has been demonstrated that regression analysis is generally robust in the presence of departures from assumptions, except for measurement errors and specification errors" (Pedhazur, 1982, p. 34). In education research, or nonexperimental research, the reliability of measurement tends to be low to moderate. According to Pedhazur, the bias in estimating regression coefficients, therefore, can be considerable (p. 34). Specification errors will be addressed in greater detail in a later section. It is sufficient to state at this point that these errors relate to the explanatory model and include:

1. Omission of important variables from the model.

2. Inclusion of irrelevant variables in the model.

3. Specifying a linear model when the actual relationships are curvilinear.

A model is an abstraction or a simplified mathematical statement of known, material relationships in the industrial or engineering form. The variables are understood to be related in the precise way in which they are defined in the model. The outcome of the analysis provides reliable data for decisions about increasing or altering the factors in the analysis. The elements of the education model, as described previously, lack the precision of the material relationships in engineering and industrial examples. The findings of the analyses, therefore, also lack the precision and usefulness for policy decisions.

Regression analysis may be used for prediction or explanation. In education, the purpose has been to explain the variation in student achievement on the basis of the inputs selected for the model. The ultimate goal is to be able to predict the effect of inputs, to calculate the cost of changes to these inputs, and to maximize student achievement for the lowest cost. The percent of the variance in the dependent variable explained by the regression model, represented by R^2, and the amount of change in the dependent variable (student achievement) for a unit change in each independent variable represented by the regression coefficient "b" associated with each independent variable (or Beta in standardized form), however, do not represent causal relationships in education (Pedazur, 1982; Cronbach, et al., 1972). The researcher is forced to recognize the speculative and uncertain nature of the findings from such an analysis.

Aside from the lack of solid, reliable data from measurable material relationships, education research is plagued by a basic problem: lack of experimental conditions. Experimental conditions allow the researcher to randomly select treatment and control groups to test hypotheses. A random sample is one that is selected by chance from a larger population. When this principle is followed, the assumption can be made that the presence of any characteristic or subgroup will be as large or small in the sample as it is in the population at large. This principle is the foundation for any conclusions the researcher can form about the larger population from the sample selected and analyzed (Howell, 1987). The education researcher must work with students, classrooms, schools, or districts that cannot be altered to meet the needs of the researcher. Even a project as huge as that conducted for the Coleman Report could not avoid the claim of systematic bias in the final result (Hanushek & Kain, 1972). Without the opportunity to randomize the samples with which a research study is conducted, the findings are diminished in power and reliability (Howell, 1987; Pedhazur, 1982).

Education production functions violate other statistical assumptions, in addition to the problem of randomization within experimental conditions. Multicollinearity occurs when independent variables show some degree of correlation. According to Pedhazur (1982), there is no consensus on the meaning of the term; this is to say that researchers cannot agree on a level of correlation between independent variables that is acceptable or not acceptable. The problem leads to difficulties in estimating the regression coefficients--the measure of the effects of the independent variables on the dependent variable. "In sum, judgment about the severity of multicollinearity in a correlation matrix remains just that: a judgment" (Pedhazur, 1982, p. 246).

Since there is no standard by which to identify the problem, it is not always easily detected. Indicators include high correlations in the correlation matrix, regression coefficients that range above "2," and wildly fluctuating regression coefficients. The latter occurrence can be summed up in this way: In a sample regression model, $Y = a + b_1X_1 + b_2X_2$, an increase in the value of "X", an independent variable, by one unit may result in the change of a "b", the regression coefficient, from a positive to a negative number. Such instability renders a regression analysis meaningless, since the interpretation of the coefficients is the basis for determining the effect of an education input on student achievement (Wolfle, 1993B; Pedhazur, 1982).

Frequently researchers actually introduce the problem of multicollinearity themselves. Since education studies must substitute available measured proxies for unmeasured variables in the empirical analysis, multiple indicators are selected to estimate, as completely as possible, the unknown variable. "When, however, multiple indicators are used in a regression analysis, they can play havoc with regression statistics" (Pedhazur, 1982, p. 242). The solution to the problem may seem simple--delete an offending proxy and proceed with the reanalysis. Unfortunately, deletion of measures to solve the multicollinearity problem can produce another serious problem--specification error.

"The general term for the description of the variables and the model is *model specification* . . . The true model is the starting point in all of our developments and the frame of reference by which to judge results. But the exact and correct formulation is not always known. The theories of the social scientist are usually not developed to the point of giving a complete model specification . . . Nor can one always expect to have the required data . . . Both of these situations, incomplete theories and incomplete data, can lead to specification errors" (Hanushek & Jackson, 1977, p. 80). Specifically, an important variable may be left out of the model, or it may be included in the model, but no satisfactory measure exists for it. Hanushek provides a useful example in Chapter Four of Education and Race. The Coleman data did not include a measure of the innate ability of students, yet, this is a variable in the model specified by Hanushek. He conducted his empirical analysis with a missing variable. Therefore, a specification bias exists in his estimated regression coefficients for this analysis.

The relationship between specification error and multicollinearity is a Scylla and Charybdis problem for the researcher. If a proxy variable is omitted from the analysis to avoid the bias resulting from multicollinearity, a specification error may result. The specification error will have the same general effect as multicollinearity--it will bias the regression coefficients. The researcher cannot have confidence that the estimated effects of the education inputs actually influence student achievement, either in the direction (plus or minus) or in the magnitude (size of the regression coefficient or "b") indicated (Pedhazur, 1982; Hanushek & Jackson, 1977).

The Problem of Missing Variables

Hanushek's conceptual models have been discussed previously. Yet, in an empirical use of the model in Education and Race (1972), he dismissed peer influences as insignificant (p. 49). Likewise, he excluded a measure of initial endowments from the model for district analyses of sixth grade students due to the lack of data (p. 55). This excellent study illustrates a typical problem with production function studies. Missing variables are sometimes ignored when data are unavailable. In such a case the findings become uninterpretable because they are manipulated solely by the availability of data and are not driven by theoretical design (Pedhazur, 1982).

The Problem of Variable Proxies

In the industrial production model, input variables are directly measurable and independent of one another. Education models cannot claim either characteristic. Multicollinearity (the result of close relationships between independent variables which confounds analysis because of the high positive correlation between proxies that may be selected to represent different variables) was discussed previously. The example cited in the section on proxies is relevant. The measures of personal income and percent of pupils eligible for free and reduced lunch may be confounding if used for community and peer characteristics respectively.

The Problem of Flawed Data: The Hanushek and Kain Critique of the Coleman Report

The Coleman Report (1966) has been a major policy influence on governmental action. The influence of this study, however, goes beyond its effect on policy. The report has had a pervasive influence on production function studies as a data source. Because the report was interpreted to mean that schools have little or no independent effect on the education of children beyond the influence of family and peers, it has been subjected to considerable analysis. The use of data from this study has been very widespread. Hence, the flaws in this data have been disseminated through many subsequent studies. The previous reference to Hanushek's (1972) violation of his own model is a consequence of this reliance upon Coleman data.

Hanushek and Kain (1972) provided a substantive critique of the Coleman data and analysis. They described problems that undermined the reliability of the report's data and analytical procedures. While the plan for data collection was enormous, many difficulties plagued the project. The student sample size of 900,000 was reduced by nonresponse to 569,000. The failure to link students to specific schools restricted the analysis to school levels only. Further reductions in usable data resulted in stratification by grade, race, region, rural/urban divisions. Nonresponse was a major problem since 41% of 1170 schools were not included in the study. The issue of systematic nonresponse was not addressed in the study, so the conclusions may be misleading. Many sensitive questions were not answered, so many questionnaire items relating to qualitative conditions were unusable. Cross checking showed many miscoded responses. The frequency of such errors raises doubt about the reliability of the Coleman survey data in general (p. 121). Other concerns involved the failure to ask questions about the quality of facilities, to collect information on per pupil expenditures, and to collect data on school organization.

The Coleman researchers never presented a general conceptual model, but they implied one throughout the text. Hanushek and Kain described an appropriate model for the Coleman analysis:

$A_{it} = g(F_i^{(t)}, P_i^{(t)}, I_i, S_i^{(t)})$ (p. 123).

A_{it} = A vector of educational achievement of the *i*th student at time *t*.

$F_i^{(t)}$ = A vector of individual and family characteristics for the *i*th student cumulative to time *t*.

$P_i^{(t)}$ = A vector of student body characteristics that represents an aggregate of socioeconomic and demographic characteristics of students in the same school as the *i*th student cumulative to time *t*.

I_i = A vector if initial endowments or inherent ability of the *i*th student.

$S_i^{(t)}$ = A vector of school inputs relevant to the experience of the *i*th student cumulative to time *t* (p. 123).

The functional model described individual achievement by the linear addition of proxies plus a random error. The equation hypothesized a consistent behavioral relationship across individuals represented by the coefficients in a regression model. This is the conceptual model Hanushek presented and modified in his subsequent works (Hanushek, 1972; Harbison & Hanushek, 1992). The statistical models employed differ notably and in systematic ways from this conceptual model, plus the divergences among them (errors in model specification) tend to bias the empirical findings toward showing inconsequential school effects.

Using Analysis of Variance (ANOVA), as the Coleman researchers did, is a useful analytical procedure if no relationship exists between independent variables, if they are indisputably independent without any correlations between them. Unfortunately, when the variables have high correlations, as they do in social research, multicollinearity makes interpretations of ANOVA "exceedingly difficult." Frequently, interaction terms become the most significant source of variance between samples and must be explained if the research is to provide meaningful information. The method of analysis of interaction variables was unusual in the Coleman Report, however. The order of variable entry in the equation was a major factor in results since interaction variance was added to the first term entered into the equation weighing its explanatory power unjustifiably and inaccurately. Family background was entered first and school inputs last. The temporal order by which a student experiences the variables may not have a necessary relationship to the way interaction terms are distributed. Indeed, the main issue is how interaction effects should be partitioned among explanatory variables. In fact, since interaction effects could not be partitioned, the Coleman Report assigned all the interactions to family background. They could as easily be assigned to school inputs. Hanushek and Kain also found that there was not enough independent variation in the school factors resulting in attenuated variance. The absence of any measure of student ability undermined the conclusion of no school effects. Since within school variance is much greater than between school variance, school inputs may vary more significantly within schools than between them. The authors of the report suggested this possibility, but the issue was not researched.

Teachers were not considered a school input in the report. Systematic departures of variables actually used in the analysis from those in the conceptual design were greatest for school inputs. For example a 12th grade student is likely to have attended many schools and experienced wide differences in educational experiences to a much greater extent than in peer and family experiences. The heterogeneity of student experiences in schools represented by vocational and college preparatory tracking obscures the variability of school experiences as reflected in mean standardized test results and school-level input aggregations.

Hanushek and Kain identified four ways in which the linear model is seriously flawed (p. 135):

1. Dimensionality problem

Logically, the effects of guidance counselors would be related to the number of students. There is no accounting for school size in terms of the impact of these or other personnel.

2. Economies of scale

The linear model does not allow for calculation of economies of scale supposedly existing in high schools.

3. Input interaction

The functional model is the same whether or not any other inputs are absent or present in any quantity.

4. Marginal value

The marginal effect of an input is the same regardless of its quantity or level of use.

According to Hanushek and Kain, the finding of no school effects is a result of the method of analysis and not the "underlying behavioral reality." Such a conclusion is "dangerous and destructive" as policy. "The extent to which minority groups are systematically discriminated against in the provision of educational inputs is still unknown. This is a serious matter since <u>the correction of input inequalities is a logical and necessary first step in insuring equality of opportunity for minorities"</u> (p. 137).

In summary, the sample size of the study was reduced by a 41% nonresponse rate. Systematic nonresponse appeared to occur, particularly on sensitive issues. Many responses were miscoded and the frequency of such errors raised doubt about the reliability of the data in general. No information was collected about per pupil expenditures, school organization, and the quality of facilities. In fact, the report found little difference between schools that were predominantly African-American or white. The data did not provide information on students by school, so that the value of a large sample was significantly reduced. Finally, the researchers stated no theoretical model with which to interpret the data.

The list of problems stemming from data and methodological issues in the Coleman Report is substantial. Several issues, however, have a generic relationship to production function studies and require elaboration. The first concern is the lack of a theory to guide the analysis (Pedhazur, 1982; Hanushek & Kain, 1972; Cain & Watts, 1970). As discussed before (see the discussion in Statistical Procedures and Problems), the use of regression analysis for explanatory purposes requires a theoretical model.

The Coleman Report, unfortunately, is not an exception in its lack of a conceptual model. Other studies, Summers and Wolf (1975, 1977) for example, not only failed to provide a model, but sought to identify important regression coefficients in a scavenger hunt approach with available data. The violation of a fundamental statistical procedure makes the findings of such studies meaningless (Pedhazur, 1982).

Another problem with the Coleman Report is the use of incremental variance partitioning to determine the relative contributions of blocks of independent proxy variables to the dependent variable, student verbal achievement. The issue is bound up with the order of entry into the equation. Where intercorrelations between independent variables occur, a shared variance exists in the relationship with the dependent variable.

(Harbison & Hanushek, 1992; Pedhazur, 1982; and Hanushek & Kain, 1972, have described the intercorrelations between variables discussed here.) The earlier position in which variables (or blocks of variables) enter the equation determines which variable or block receives "credit" for the shared variance. "Because the background differences are prior to school influence, and shape the child before he reaches school, they will, to the extent we have succeeded in measuring them, be controlled when examining the effect of school factors. This means that the achievement differences among schools that are due only to differences in student input can be in part controlled, to allow for more accurate examination of the apparent effects of differences in school or teacher factors themselves" (Coleman, et al., 1965, p. 298). The temporal criteria for entering student background variables first and school variables last guaranteed that school factors would demonstrate a smaller influence than background variables regardless of the underlying reality in school achievement. Since the correlation between school quality and student background characteristics tends to be quite high, the analysis actually partialed out school effects (Pedhazur, 1982; Bowles & Levin, 1968). In fact, since the R^2, or the proportion of variance explained by a regression equation, is sample specific and may vary from one sample to another regardless of the underlying reality, it is questionable to base inferences about populations on the statistic (Pedhazur, 1982). Added to this limitation is the statistical impossibility of apportioning shared variance between student background variables and school inputs, or any other combination of variables. The probability that serious misinterpretations will occur about the effects of any of the independent variables on student achievement is clear.

Pedhazur takes explicit exception to the notion that shared variance represents interaction effects. Critics (note the summary of Hanushek and Kain previously discussed) as well as the Coleman researchers accept the premise that shared variance represents interaction effects. The total effect of correlated variables on the dependent variable represent a sum of direct and indirect effects. These indirect effects represent correlations of exogenous variables whose variability is not explained by the model. Endogenous variables are those whose variation is explained theoretically entirely by the model. The unexplained variance between correlated exogenous variables such as school and student background effects lies beyond the conceptual model, much less the empirical one. Therefore, the assumption of interaction between the exogenous variables is unwarranted. In summary, "because the elements obtained in incremental partitioning of variance reflect different types of effects, it is *not* valid to compare them for the purpose of determining their relative effects on a dependent variable" (Pedhazur, 1982, p. 188).

The Unit of Analysis and Cross-Level Inferences

Cross-level inferences occur when the researcher uses data aggregated at one level to reach conclusions about units at another level (Pedhazur, 1982). Levels of data for education production studies have been aggregated at the district level, school level, classroom level, and individual student level. While inferences are sometimes made from the individual to a larger aggregate, inferences made from the larger aggregate to the individual has been more common in education (Roberts & Burstein, 1980). The problem is particularly acute in the analysis of high school data. When the mean of teacher experience for the entire school is correlated to the mean achievement scores of individual students, all of whom had only a small percentage of the teachers, the results should be viewed with skepticism. This problem increases with the size of the high school, because a diminishing percentage of students study under any given teacher (Glasman & Binianimov, 1981).

If each student's achievement in an analysis could be matched with specific data about her teachers, family, peers, and school expenditures, the individual unit of analysis could be preserved across all levels of input. Unfortunately, information of this specificity has yet to be collected. Aggregate data, collected for other school, census, or tax purposes have been used.

Glasman and Biniaminov (1981) described several studies that mixed levels of data aggregation for a single analysis. The Coleman Report is the most widely known, but twelve other studies were cited. Researchers who use the individual student or the classroom as the level of analysis are forced to make cross-level inferences if they employ a model that incorporates the elements assumed to influence achievement (see the previous section on the education model).

Inferences from correlations between cross-level data have been shown to be seriously fallacious (Robinson, 1950; Bidwell & Kasarda, 1980; Pedhazur, 1982). With data correlated at both the individual level and a higher aggregate level, Robinson demonstrated in several examples that inferences from cross-level correlations could differ not only in magnitude but by sign as well. This problem has become known as the ecological fallacy.

While it is possible for an ecological fallacy to be committed when correlations are used, and not when regression is used, differences ultimately relate to the way groups are formed. The issue of randomization lies at the heart of the problem. Education research has not been based on experimental conditions that allow for randomization of sample selection or assignment. Only with randomized conditions are unbiased regression coefficients a safe assumption. "The main problem is that when intact groups are studied it is very difficult, often impossible, to unravel the processes by which the groups were formed. When, under such circumstances, data are available on the group level only it is generally not possible to tell the direction and the magnitude of the bias that results from inferences made about individuals" (Pedhazur, 1982, p. 539).

Another problem to consider in the level of aggregation is the proportion of variance explained by the model, R^2. When the individual is the unit of analysis, the proportion of variance explained directly relates to the total variance. When aggregates are analyzed at higher levels, such as classrooms or schools, the R^2 explains only the proportion of variance between groups. The within group variance remains unexplained because of the level of aggregation. Pedhazur (1982) uses the example of the between group variance that explains only .10 of the total variance. Yet, the R^2 appears to be quite high in explaining .80 of this variance. The high R^2 actually explains only .80 of the .10. The aggregation of information above the individual student level results in the explanation of a very small percentage of the total variance in student performance.

Mixing data from different levels can introduce significant errors into a study and undermine the findings of the analysis, much less any conclusions that might be offered. This problem is a persistent concern in the interpretation of education production functions.

Related to the ecological issue are contextual effects. A contextual effect is the net effect of a group variable after controlling for the effect of the same variable on the individual level (Pedhazur, 1982). A methodological process exists to analyze group effects when proper controls are included. This procedure should not be confused with the problems of analysis based on data aggregated at different levels.

The Variety of Perspectives in Empirical Applications

A survey of different perspectives on the education production function outlines the framework of existing approaches. The following brief examples are presented in a chronological sequence. The Equality of Educational Opportunity Report (called the Coleman Report, 1966) has been the most influential of the production function analyses in education. Since the project was funded by Section 402 of the Civil Rights Act of 1964, it remains one of the most comprehensive collections of data used in these analyses. Its influence has been discussed in previous sections of this paper. Criticisms of the methodology, data quality, and conclusions were also described.

Burkhead (1967) used Project Talent data to compare small-community high school achievement to that of Chicago and Atlanta. He described the education formula as an exploration, not a true production function. Unlike the industrial model, he recommended changing factor combinations to find the best configuration for increasing education output. This study included a value-added approach in the Chicago and Atlanta studies. He described three levels of resource use: acceleration, perpetuation, and amelioration. Acceleration referred to the practice in Atlanta in the early 1960s of investing more money in wealthier, higher achieving white schools. The perpetuation design involved equal expenditure across all students. This maintained the existing condition in achievement reflected in the Chicago schools of the period. Amelioration required greater investment in schools for lower income and lower achieving students. His study demonstrated the importance of family income in student achievement. To improve student performance, he found it necessary to break the linkage of educational inputs to community income levels. If family income could not be changed, improvement in school outputs required dramatic increases in inputs or significant changes in resource combinations. This proposition could lead to the assumption of a unidentified threshold effect for breaking the link between family income and student success, as well as experimentation with resource combinations in educational production.

Hanushek (1972) emphasized the importance of production function studies in policy decisions. His central concern was the education of minorities and the ability of education to cure the condition of minorities in income, jobs, and life expectancy. Hanushek is careful to identify the problematic nature of conclusions from his analyses.

Because of the complexity of high school production functions and the difficulty of assessing cumulative factors in student performance, Hanushek examined elementary schools. He collected data at the individual school level in one of his three analyses. The other two studies were based on data from the Coleman Report.

Hanushek assumes that a public institution is inefficient because it does not operate in an openly competitive market. Incentives do not exist for the efficient maximization of resource use or educator performance. The production study could identify, with greater precision than existing methods, where efficiencies could be achieved. Hanushek defined efficiency as consisting of two elements: knowledge of the relationship between inputs of the educational process and a decision rule to connect costs of various inputs to the process with their educational outputs. Armed with this knowledge, policy makers could mix resources in ways that would be efficient and productive for student achievement. He did three studies that suggested areas of inefficiency in the purchase of teacher experience and additional education. Schools were incapable of curing the condition of minorities without changing the pattern of expenditures.

Rossmiller (1986) evaluated public elementary school classrooms in relation to equity and efficiency. Analyzing data at the classroom level enabled he and Frohreich to examine the achievement of students who directly received specific resources at precise costs. Time utilization, school expenditures, home environment, and teacher characteristics and attitudes were all carefully studied. Money and time correlated negatively with achievement because of the greater application of these resources to assist lower achieving students. This suggested a trade-off between economic efficiency and efforts to ameliorate student achievement deficits.

Thompson and Correa (1989) researched a specific cohort of students in a private elementary school setting. They rated the cumulative effect (over three years) of certain input variables on student achievement. Using Glasman and Biniaminov (1983) as a reference, they found differential effects of these variables on private and public school students. As in public school analyses, teacher and school variables played a minor role in the prediction of variance in student achievement. Student academic ability was more important than school inputs. Teacher fluency, academic degree, and annual salary correlated with higher mathematics achievement in contrast with both reading and mathematics in public schools. Class and school size correlated negatively with private school achievement, but positively in some public school studies. The authors concluded that "caution should be used when applying the results of effective school research conducted in the public schools directly to the private schools" (p. 406). The results of this private school study, where there is a competitive market, did not differ in any significant way from the results of public school analyses.

Ferguson (1991) used an unusually large and complete data set from 900 districts in Texas that provided information on a student population five times that of the Coleman Report. Using a district-level analysis, Ferguson examined the determinants of student test scores, factors that influence which districts attract the most effective teachers, and how and why money matters in student achievement.

According to Ferguson's findings, money is important in producing higher student test scores when it purchases teachers with strong literacy skills, reduces class size to eighteen students per teacher, retains experienced teachers, and increases the number of teachers with advanced degrees. In addition to equalizing funding per pupils, ameliorative programs for low SES districts could include state subsidized higher pay for teachers relative to higher SES districts. Good teachers are attracted to higher SES districts and the salaries they pay. Offering salary subsidies is a market decision that would encourage higher quality teachers to accept the challenge of teaching in difficult conditions. Possibly, this could begin to break the link Burkhead found between family income and student achievement. This study uses other recent research to undermine the basic conclusion of the Coleman Report and the findings of less data-rich studies about the impact of teacher quality on student achievement.

CHAPTER IV

The studies described in this chapter identified positive and significant findings for the direct or indirect influence of financial expenditures on student achievement. They were selected from journals, dissertations, and ERIC documents that focused on the K-12 education environment from the era of the Coleman Report to the present. This sample of studies utilized a range of statistical methodologies from simple correlation to hierarchical linear modelling. The evaluation of the effects of independent variable proxies on specific achievement measures are examined by disaggregating the analysis by sample size, level of aggregation, grade level, location, effect, significance, and skills test. A discussion of each study's relationship to a conceptual model of the education production function provide a more secure foundation for understanding what education production studies can tell us about student achievement and suggest improvements in the quality of such studies.

Pelton, see Table 1, tested the correlation between perceptions of educational quality as measured by the Educational Characteristics Criterion (ECC), selected expenditure variables, and student achievement. The Educational Characteristics Criterion was a questionnaire developed by Professor Herbert Rudman designed to measure the quality of education through the perceptions of knowledgeable observers, principals and teachers in this study. Contemporary with the Coleman Report, the questionnaire collected responses to 55 questions grouped in seven categories relating to: students' level of knowledge and attitudes, community attitudes, curriculum, use of

Table 1. Study Characteristics of Pelton

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per pupil expenditure	district	composite score	district	(+) S*	19	6	United States
state equalized valuation (ability to pay)	district	composite score	district	(+) NS	19	6	United States
tax rate in mills (effort)	district	composite score	district	(+) S#	19	6	United States
average daily membership	district	composite score	district	(+) NS	19	6	United States
teacher TQS	district	composite score	district	(+) NS	19	6	United States
administrator TQS	district	composite score	district	(+) NS	19	6	United States

Other Comments:
Scores for the 1964 Stanford Achievement Test from 19 districts in 11 states were used in the analysis. The findings reported in this chart are based on correlation coefficients. The level of significance was set at .10, indicated by S#. Total Quality Scores (TQS) were the results of the Educational Characteristics Criterion (ECC) questionnaire administered to 1486 teachers and 131 administrators in the sample of 19 districts.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

facilities, sociocultural composition of the community, administration and supervision, and teachers and teaching methods.

Pelton found that teachers and principals, generally, do perceive the characteristics of quality education in the same way. Differences emerged more clearly in characteristics related to the teaching process. Administrator perceptions of quality correlated more highly with student achievement than teachers. Administrators perceptions also correlated highly with per pupil expenditures. Cost factors, per pupil expenditures and millage rates, were statistically significant and positive.

In relationship to the conceptual model, this study divided data into perceptual and measured sources. There are data on family background, community, peers, student ability, and school characteristics. Aside from test scores and financial data, however, the remainder were perceptual data and entered the assessment through the Total Quality Scores (TQS) of principals and teachers. The design did not provide a separate correlation of the questionnaire categories with student achievement. Moreover, the perceptions of principals and teachers as measured by the TQS were not statistically significant in relation to student achievement scores.

The ECC broadly resembles contemporary effective school questionnaires such as those used by New York state. The survival of this approach to specific school improvement is an alternative method for acquiring data on the quality of schools in contrast to the reliance on limited, existing quantitative data. The Coleman Report, itself, used questionnaire data for part of its assessment, and subsequent education production function studies have sometimes incorporated such perceptual data into an analysis. This methodology undermines the objective economics-based origin of the production function analysis. For an analysis beyond the individual building level, it seriously erodes the researcher's ability to claim scientific objectivity in a study already lacking random selection and limited by sample size and representativeness. In historical context, the study has merit as a step toward the use of perceptual data in a quantitative context to identify educational program needs. It cannot explain the relationships between variables and student achievement.

Booms, Table 2, attempted to identify empirical estimates of secondary education production functions. Project Talent data collected from a 5% sample of U.S. high schools (grades 9-12) in 1960 was the primary information source. In addition, the National Inventory of School Facilities and Personnel collected by the Department of Health, Education, and Welfare (H.E.W.) in 1962 provided information on capital investments in building facilities. The author used "real" proxies for variables rather than financial measures on the premise that costs will vary, but real variables will indicate an enduring relationship. The author used stepwise regression to identify statistically significant proxies from the variable categories of labor, capital, technology, organization, scope, and student-societal factors. The significant variables were used in a linear regression equation and a multiplicative model. In each case the equations were analyzed with and without capital investment proxies.

Using twelfth grade social studies scores as the dependent variable, Booms found homework, number of teachers with Masters degrees, the number of courses offered in school, ninth grade test scores, an SES index, number of rooms in the school, and the

Table 2. Study Characteristics of Booms

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
homework	school	social studies test R-105 from Project Talent	school	(+) S**	648	12	United States
number of teachers with MA	school	social studies test R-105 from Project Talent	school	(+) S**	648	12	United States
number of courses offered in school	school	social studies test R-105 from Project Talent	school	(+) S**	648	12	United States
9th grade scores	school	social studies test R-105 from Project Talent	school	(+) S**	648	12	United States
SES index	school	social studies test R-105 from Project Talent	school	(+) S**	648	12	United States
number of rooms in school	school	social studies test R-105 from Project Talent	school	(+) S**	648	12	United States

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 2 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
number of makeshift rooms in school	school	social studies test R-105 from Project Talent	school	(-) S*	648	12	United States

Other Comments:
The findings reported in this chart are based on a linear regression equation identified as the additive gross with capital variables. This equation represents a summary of variables selected by stepwise elimination in categories identified by the author (from Tinbergen) as appropriate to the production of educational achievement.
The use of a standardized social studies test as the output measure in the 12th grade is unusual.
Project Talent data come from a 5% sample of U.S. high schools (grades 9-12) taken in 1960. In addition, the author used information from the National Inventory of School Facilities and Personnel compiled from 1962 data for the Department of Health, Education, and Welfare. This reduced the total sample size for the analysis using capital investment variables, but it allowed for incorporation of facilities information into the study.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

number of makeshift rooms in school as statistically significant independent variables in the final regression model. Although the findings were similar for both equations, Booms concluded that the linear model was a better fit than the multiplicative one based primarily on its larger R^2.

Despite the author's use of "real" proxies, there are financial implications in his analysis. Without cost data for the variables, there is no possibility for comparing the value of choosing one variable over the other in terms of expenditures. This problem has consistently undermined the value of any policy application for education production function studies. The finding that improved teacher education, permanent facilities, and course offerings have positive and significant regression coefficients for student achievement suggests that expenditures for these items contributed to student achievement in 1960.

In this work the author proposed the possibility that the estimation of a high school production function could be a single equation with applicability on a national level. Subsequent research has cast serious doubt upon that proposition. Since data from Project Talent and the H.E.W. facilities study are separated by two years, it does not seem valid to merge it for a cross-sectional analysis. The use of school level data from Project Talent, however, offered the opportunity for a better estimate of an education production function because it more closely related to the actual educational experience of the students who were analyzed. Many studies have used district level data. The further data are aggregated from the student level the less likely meaningful production estimates can be obtained.

Dunnell, Table 3, used incremental partition of variance and stepwise multiple regression to determine the relationship between student achievement and socioeconomic (SES) characteristics of a community, financial resources, cost of the educational program, and educational treatments in 44 suburban school districts in Cook County, Illinois. For each of these four factors, one to fourteen proxy variables were selected from the available data to be analyzed. From a total of 119 elementary districts, a sample of 44 districts was selected for the 1967-68 school year because they used the Stanford Achievement Test, 1964 edition, to evaluate both fourth and seventh grade students. The researcher collected test data from the superintendents of the districts, expenditure information from the state education department and local districts, and socioeconomic data from the 1960 census.

The author found positive and significant effects for financial, demographic, and school inputs from the stepwise regression as shown in Table 3. In particular, per pupil operating expenditures and assessed valuation per pupil were positive and significant. Among the SES proxy variables median value of homes, median years of schooling, and median income of families demonstrated a very strong positive and significant effect. In the incremental partition of variance, however, only the SES factor was significant. The latter finding led the author to conclude that SES is the major determinant of student achievement. Among the proxies that can be purchased or manipulated by administrators, lower class sizes and the number of special area teachers in art, library, vocal and instrumental music, physical education, and reading were important to improved achievement.

Table 3. Study Characteristics of Dunnell

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
median value of homes	county	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
special area teachers per 1,000 pupils	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
percent ADA to enrollment	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
median income of families	county	composite	district	(+) S* (+) NS	44 districts	4 7	Cook County, Illinois
median years of schooling	county	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
percent of nonwhite pupils	district	composite	district	(-) S** (-) S**	44 districts	4 7	Cook County, Illinois
beginning Master's salary	district	composite	district	(+) NS (+) S**	44 districts	4 7	Cook County, Illinois

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 3 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per pupil operating expenditure	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
assessed valuation per pupil	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
beginning Bachelor salary schedule	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
percent of professional and managerial workers	county	composite	district	(+) S** (+) NS	44 districts	4 7	Cook County, Illinois
percent with Master's degree	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 3 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
central office administrators per 1,000 pupils	district	composite	district	(+) S** (+) NS	44 districts	4 7	Cook County, Illinois
enrollment	district	composite	district	(+) S** (+) S**	44 districts	4 7	Cook County, Illinois
library volumes per pupil	district	composite	district	(+) NS (+) S**	44 districts	4 7	Cook County, Illinois
pupils per building	district	composite	district	(-) NS (-) S**	44 districts	4 7	Cook County, Illinois
income over $10,000	county	composite	district	(-) NS (+) S**	44 districts	4 7	Cook County, Illinois
population of community	county	composite	district	(-) NS (-) S**	44 districts	4 7	Cook County, Illinois
years to reach maximum salary	district	composite	district	(+) NS (+) S**	44 districts	4 7	Cook County, Illinois
average class size	district	composite	district	(-) NS (-) S**	44 districts	4 7	Cook County, Illinois

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 3 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent of male teachers	district	composite	district	(+) NS (-) S**	44 districts	4 7	Cook County, Illinois

Comments:
All results in this chart are based on correlation coefficients. The Office of the Superintendent of Public Instruction for Illinois, and the Cook County Superintendent provided financial data. The 44 district superintendents provided educational variable achievement test data for 1967-68. Socioeconomic data were compiled from 1960 Bureau of the Census. Estimates were made when school district boundaries did not coincide with census tract maps.
Only elementary (K-7) school districts in Cook County (excluding Chicago) that used the Stanford Achievement Test for both grades 4 and 7 were selected to participate in the study.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

This study was among the early analyses to follow the Coleman Report. It incorporated elements of the conceptual model that emerged during this period by utilizing the linear regression methodology and selecting proxies to represent variables hypothesized to influence student achievement. Stepwise regression, however, is useful for prediction. It is an inappropriate procedure for explanatory investigations such as those undertaken in this analysis. Likewise, "controlling variables through partialing without regard to the theoretical considerations about the pattern of relations among them may amount to a distortion of reality and result in misleading or meaningless results" (Pedhazur, p. 110, 1982). Missing from the study were proxies for student ability or previous achievement. Socioeconomic (SES) variables were not directly related to the pupils who were tested. SES proxies were taken from 1960 census data for the county as a whole, not specific school districts. Adjusting county data to reflect the percentage of the population of a school district can only result in correct variable specification if a district is a microcosm of the county as a whole. In a limited sample, if the range is great--as indicated in this study by property valuation, then taking a percentage of census data dramatically skewed SES measures. If the range were small, then the sample is insufficiently dichotomous producing a truncated variance effect. Likewise, the use of 1960 data to represent a 1967-68 group ignored changes in the county as a whole, much less selected school districts within the county. SES proxies represented neither an accurate picture of the pupils' family backgrounds, peer characteristics in the schools, nor community characteristics relevant to the group whose performance was analyzed.

The expenditure data represented a five year average. Such an approach may have merit in representing the resources applied to the given group of students under study. Most studies use cross-sectional financial data to parallel that of the other proxies. The time-frame within which financial resources have an effect on students remains unclear since expenditures occur at varying times, in varying amounts, and through varying resources.

Administration of the Stanford Achievement Tests in the districts under study ranged over eight months. The use of an average score as the dependent variable distorted the performance of any given group. However, the use of the scores as administered also would have introduced an obvious distortion, suggesting a performance gap that was actually determined by the range in testing dates. Another problem with the dependent variable was the use of grade equivalent (GE) scores. They are not interval level data. Standardized scores would have provided a higher quality analysis.

While the study was thorough and carefully structured, the problem of data availability undermined the usefulness of the analysis and findings. The problem continues today in education production function studies. While some of the difficulties, such as the uniformity of testing dates for a given sample, have largely been overcome in more recent analyses, modifying census demographic data for a political subdivision to a school district within that division is a continuing dilemma.

Stinson and Krahmer, Table 4, suggested that per teacher expenditures are a more efficient method of allocating resources than per pupil measures. Using composite achievement scores for 9th and 11th grade pupils from the 1965 administration of the

Table 4. Study Characteristics of Stinson and Krahmer

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per pupil instructional cost	group of high schools	composite	group of high schools	(+) NS (+) NS	66 high schools	9 11	North Dakota
per pupil operating cost less transportation	group of high schools	composite	group of high schools	(+) NS (+) NS	66 high schools	9 11	North Dakota
per pupil operating cost	group of high schools	composite	group of high schools	(+) NS (-) NS	66 high schools	9 11	North Dakota
per pupil total operating cost less transportation	group of high schools	composite	group of high schools	(-) NS (+) NS	66 high schools	9 11	North Dakota
per pupil total cost	group of high schools	composite	group of high schools	(-) NS (-) NS	66 high schools	9 11	North Dakota
per teacher instructional cost	group of high schools	composite	group of high schools	(+) S** (+) S**	66 high schools	9 11	North Dakota
per teacher operating cost less transportation	group of high schools	composite	group of high schools	(+) S** (+) S**	66 high schools	9 11	North Dakota

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 4 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per teacher operating cost	group of high schools	composite	group of high schools	(+) NS (+) NS	66 high schools	9 11	North Dakota
per teacher total operating cost	group of high schools	composite	group of high schools	(+) NS (+) NS	66 high schools	9 11	North Dakota
per teacher total operating cost less transportation	group of high schools	composite	group of high schools	(+) S** (+) S**	66 high schools	9 11	North Dakota
per teacher total cost	group of high schools	composite	group of high schools	(+) NS (+) NS	66 high schools	9 11	North Dakota

Comments:
Composite scores from the 1965 results for the Iowa Test of Educational Development, a standardized achievement test, were used for the analysis.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Iowa Test of Educational Development and financial data from North Dakota school districts, the authors tested the thesis. They found statistically significant correlations between per teacher expenditures and composite pupil achievement test scores using instructional costs, operating costs less pupil transportation, and total costs less pupil transportation as independent variables. They found no significant correlations between per pupil spending and achievement using these same measures. Rural high schools were divided into three size categories for the analysis. Differences between schools based on size and grade were few and insignificant.

Stinson and Krahmer used correlation analysis only. A theoretical model using demographic data aside from school size was not considered. The findings suggested a different approach to examining the allocation of resources that would be difficult to duplicate: allocation on a per teacher basis. The work of Frohreich and Rossmiller (1986) provides a more sophisticated approach to allocation at the classroom level. They found no statistical significance due in part to the focus of teacher attention on the least able pupils. The 1965 data are certainly time-bound since they precede the advent of special education. The higher per pupil and per teacher resource allocation to contemporary handicapped programs would inhibit any correlation between achievement and per teacher expenditures. An analysis that excludes them from any evaluation involving expenditures and achievement for schools and classrooms to which they belong distorts the reality of the current educational experience. Exceptional pupils may be given non-standardized tests, but the investment of teacher time and resources is a zero sum game. Teachers may spend less time and resources with those who are evaluated on achievement tests as a result of their effort to meet the greater needs of exceptional pupils. This study is useful in suggesting to us the resulting distortions introduced into contemporary studies of expenditures and achievement.

Jewell, Table 5, analyzed fiscally related variables and measures of student achievement to determine the nature of the relationship between them. He used data from twenty public school districts in grades four, eight, and eleven representing a geographic sampling of Kentucky, excluding large cities. The sample was determined by the need to use uniform test and fiscal data for 1969-70. Data were collected from the participating school districts and the Kentucky Department of Education. Although the author reported findings for the relationship using average district achievement scores, the analysis was characterized by an effort to use a value-added approach in measuring achievement. Jewell used linear regression to determine adjusted achievement scores from the Otis-Lennon Mental Ability Tests and the Stanford Achievement Test Battery. Correlation and stepwise multiple regression data were reported in the analysis.

The author concluded that there was a positive relationship between per pupil expenditures and the quality of education. He found the highest relationship between achievement measures and the percent of total expenditures provided from local sources. High positive relationships existed for annual expenses per pupil in average daily attendance, percent of district supplement for instructional salaries, cost per pupil for instructional materials, local financial index, percent of attendance, cost per pupil for administration, and cost per pupil for instruction. Negative correlations were found between achievement and percent of revenue from state sources and percent of revenue

Table 5. Study Characteristics of Jewell

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent of revenue from local sources	district	verbal	district	(+) S** (+) NS	20 districts	4, 11 8	Kentucky
		quantitative		(+) S**		4, 8, 11	
per pupil expenses to A.D.A.	district	verbal	district	(+) S**	20 districts	4, 8, 11	Kentucky
		quantitative		(+) S**		4, 8, 11	
per pupil instructional materials costs	district	verbal	district	(+) S** (+) S*	20 districts	4, 8 11	Kentucky
		quantitative		(+) S** (+) NS		4, 8 11	
local financial index	district	verbal	district	(+) S** (+) NS	20 districts	4, 11 8	Kentucky
		quantitative		(+) S** (+) NS (+) S*		4 8 11	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 5 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percentage attendance	district	verbal quantitative	district	(+) S** (+) S** (+) NS	20 districts	4, 8, 11 4, 11 8	Kentucky
percent of district supplement for instructional salaries	district	verbal quantitative	district	(+) S** (+) S** (+) NS	20 districts	4, 8, 11 11 4, 8	Kentucky
per pupil administrative costs	district	verbal quantitative	district	(+) S** (+) NS (+) S**	20 districts	4, 8 11 4, 8, 11	Kentucky
percent of instructional staff with less than rank III	district	verbal quantitative	district	(+) S** (+) NS (+) S** (+) S*	20 districts	4, 11 8 4, 11 8	Kentucky

Comments:
The relationship between average verbal and quantitative scores from the Stanford Achievement Test Battery and selected independent variables resulting from the predictive stepwise multiple regression are displayed in this table.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

from federal sources. Jewell found little relationship between achievement measures and teacher-pupil ratio, teacher certification, or annual average salary for classroom teachers. More variables had significant predictive value at each grade level for average achievement than for average adjusted achievement. From this finding, Jewell concluded that interest, background, environment, motivation, and other variables should be used for predicting achievement scores. Pupil performance in ability and achievement was highly consistent for the highest and lowest performing districts across all grade levels and tests. Exceptions to this consistency between grade level performance within some districts led Jewell to suggest that errors existed in the data.

The Jewell study was placed in the tradition of cost-benefit analysis. In this context the study failed to include several variables from the production function conceptual model. Interestingly, these are variables that Jewell recommended for use in future studies. Proxies for family, community, and peer variables were limited or missing from the analysis. The use of value-added achievement measures provided, however, the one variable from the conceptual model most often lacking in other studies. While Jewell made reference to the use of data for 17,170 pupils, the sample size is reduced to 20 due to the level of data aggregation at the district level. Such a limited sample size further handicapped by a lack of randomization provides no basis for generalization of the findings to a larger population.

The California State Department of Education, Table 6, used 15 indicators of school output quality and 18 indicators of input quality to perform correlation and regression analyses of every district in the state. These data were used for comparisons

Table 6. Study Characteristics of California State Department of Education

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
minimum teacher salary	district (unified)	reading	district	(+) S** (+) NS	235 districts	1,2,3,6 12	California
		language		(+) S** (+) NS		6 12	
		spelling		(+) S** (+) S*		6 12	
		math		(+) S**		6,12	
maximum teacher salary	district (unified)	reading	district	(+) S** (+) NS	235 districts	1,2,3,6 12	California
		language		(+) S**		6,12	
		spelling		(+) S**		6,12	
		math		(+) S**		6,12	

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
median teacher salary	district (unified)	reading	district	(+) S** (+) S*	235 districts	1,2,3,6 12	California
		language		(+) S**		6,12	
		spelling		(+) S**		6,12	
		math		(+) S**		6,12	
average class size grades 1-3	district (unified)	reading	district	(-) NS (+) NS (-) S*	235 districts	1,2,6, 12 3	California
		language		(-) NS (+) NS		6 12	
		spelling		(-) NS		6,12	
		math		(-) NS		6,12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
pupil teacher ratio grades 4-8	district (unified)	reading	district	(-) NS (-) S*	235 districts	1,2 3,6	California
		language		(-) NS		6,12	
		spelling		(-) NS (+) NS		6 12	
		math		(-) S* (-) NS		6 12	
non-teaching personnel	district (unified)	reading	district	(+) S** (+) S* (+) NS	235 districts	1,3 6 2,12	California
		language		(+) S** (+) S*		12 6	
		spelling		(+) S**		6,12	
		math		(+) NS		6,12	

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
general fund tax rate	district (unified)	reading	district	(+) NS (-) NS	235 districts	1,2,6 3,12	California
		language		(+) NS		6,12	
		spelling		(+) NS		6,12	
		math		(+) NS		6,12	
general purpose tax rate	district (unified)	reading	district	(+) NS (+) S*	235 districts	1,2,3,12 6	California
		language		(+) S* (+) NS		6 12	
		spelling		(+) S* (+) NS		6 12	
		math		(+) NS		6,12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
assessed valuation/ ADA	district (unified)	reading	district	(+) NS (+) S** (+) S*	235 districts	1,2,12 3 6	California
		language		(+) NS		6,12	
		spelling		(+) NS (-) NS		6 12	
		math		(+) NS (+) S*		6 12	
minority enrollment	district (unified)	reading	district	(-) S**	235 districts	1,2,3,6,12	California
		language		(-) S**		1,2,3,6,12	
		spelling		(-) S**		6,12	
		math		(-) S**		6,12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
index of family poverty	district (unified)	reading	district	(-) S**	235 districts	1,2,3,6,12	California
		language		(-) S**		6,12	
		spelling		(-) S**		6,12	
		math		(-) S**		6,12	
scholastic ability grade 6	district (unified)	reading	district	(+) S**	235 districts	1,2,3,6,12	California
		language		(+) S**		6,12	
		spelling		(+) S**		6,12	
		math		(+) S**		6,12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
scholastic ability grade 12	district (unified)	reading	district	(+) S**	235 districts	1,2,3,6,12	California
		language		(+) S**		6,12	
		spelling		(+) S**		6,12	
		math		(+) S**		6,12	
pupil mobility grades 1-8	district (unified)	reading	district	(-) S** (-) NS	235 districts	1,2,3,6 12	California
		language		(-) S**		6,12	
		spelling		(-) S** (-) NS		6 12	
		math		(-) S**		6,12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
pupil mobility grades 9-12	district (unified)	reading	district	(-) S**	235 districts	1,2,3,6,12	California
		language		(-) S**		6,12	
		spelling		(-) S** (-) S*		6 12	
		math		(-) S**		6,12	
rate of staff turnover	district (unified)	reading	district	(-) S** (-) NS	235 districts	1,2,3,6 12	California
		language		(-) S**		6,12	
		spelling		(-) S** (-) S*		6 12	
		math		(-) S* (-) S**		6 12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
expenditures per ADA	district (unified)	reading	district	(+) S** (-) NS	235 districts	1,2,3,6 12	California
		language		(+) S** (+) NS		6 12	
		spelling		(+) S** (+) NS		6 12	
		math		(+) S** (+) S*		6 12	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 6 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
regular ADA grades 1-12	district (unified)	reading	district	(+) S* (+) S** (+) NS	235 districts	1,6 2 3,12	California
		language		(+) S** (+) NS		6 12	
		spelling		(+) S**		6,12	
		math		(+) S** (+) S*		6 12	

Comments:
All results in this chart are based on correlation coefficients. The California Department of Education reported these scores for all districts in the state for 1969-70. Only unified districts (those that include grades 1-12) are reported here. The Cooperative Primary Reading Test was used for grade 1; the Stanford Reading Test was used for grades 2 and 3; the Comprehensive Test of Basic Skills was used for grade 6; and the Iowa Test of Educational Development was used for grade 12.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

with statewide averages. Districts were aggregated into 253 unified, 723 elementary, and 120 high school groups that were subdivided into non-uniform size ranges.

The analysis determined that high pupil achievement test scores were most related to high pupil scholastic ability scores and low rates of family poverty. In addition, high achievement scores were associated with high teachers' salaries, larger districts, high expenditures per pupil, high proportions of nonteaching certificated personnel, low percentages of minority group pupils, low rates of pupil mobility, low rates of staff turnover, small class sizes, and low pupil-teacher ratios.

The California report contains several unique features. The use of Lorge-Thorndike Intelligence Tests--Verbal Battery scores as a measure of initial pupil endowments or ability renders the analysis more conceptually complete than many production function studies. Although no explanatory model was identified in the report, regression analyses were used to predict performance levels for each district. The residuals between prediction and actual achievement test performance were charted according to below, within, and above range accomplishment. In addition, the department used a broad selection of district level financial measures based on the proposition that "it is necessary to consider the ability and the willingness of school districts to provide for a quality education program" in order to properly interpret the results of achievement tests.

Rumbaugh, et al., Table 7, found that per pupil instructional expenditures demonstrated a statistically significant and positive correlation with pupil achievement at grades four and seven for over 500 districts in Michigan. Theorizing that district size

Table 7. Study Characteristics of Rumbaugh, Donovan, Huyser, and Schooley

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per pupil instructional expenditure	all districts	composite MEAP scores	all districts	(+) S* (+) S*	504 districts	4 7	Michigan
per pupil instructional expenditure	districts larger than 10,000 pupils	composite MEAP scores	districts larger than 10,000 pupils	(+) S* (+) S*	33 districts	4 7	Michigan
per pupil instructional expenditure	districts with 2,000 to 9,999 pupils	composite MEAP scores	districts with 2,000 to 9,999 pupils	(+) NS (+) NS	224 districts	4 7	Michigan
per pupil instructional expenditure	districts smaller than 2,000 pupils	composite MEAP scores	districts smaller than 2,000 pupils	(+) S* (+) S*	257 districts	4 7	Michigan

Other Comments:
The Michigan Educational Assessment Program (MEAP) composite scores were compiled by averaging individual pupil standard scores on reading, mechanics of written English, and mathematics across districts for 1970-71 results. The enrollment data were based on information reported for the 1969-70 academic year. Correlation coefficients between K-12 instructional expense per pupil and composite achievement were reported for grades four and seven. The highest R^2 achieved was .25 for districts with 10,000 pupils and above.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

might influence the relationship, the authors grouped districts by enrollment according to the following criteria: greater than or equal to 10,000 pupils; 2,000 to 9,999 pupils; and less than 2,000 pupils. While larger and smaller districts continued to show significant positive correlations, the middle range of districts with membership between 2,000 and 9,999 did not show a statistically significant relationship even though the group had higher composite achievement scores and lower instructional expenses than larger districts. The authors concluded that money does make a difference in pupil achievement, although it is not the only indicator of quality.

The authors hypothesized a limited relationship based on only three variables. Although the relationship was positive and significant, the overall correlation between per pupil expenditure and achievement accounted for only two percent of the variance. An additional feature of this study was a survey of pre-Coleman cost-quality studies that showed a positive relationship between expenditures and achievement.

Using data from the Coleman Report to examine pupil achievement in verbal, non-verbal, mathematics, reading, and general information measures, Boardman et al., Table 8, analyzed a two stage least squares (2SLS) model to allow for reciprocal influences between independent endogenous variables.

For school characteristics that are purchased, the authors found significant positive effects for teacher verbal achievement, the number of teachers per pupil, and a curvilinear effect for teachers' experience. Teachers show a declining effect during the first few years on the job (assumed to result from a decline in enthusiasm), but experience begins to have an increasing impact over time. The age of the facility had

Table 8. Study Characteristics of Boardman, Davis, and Sanday

Input Variables		Output Variables		Direction of Effect Reduced Form	Direction of Effect Structural Form	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation					
teacher's average years of experience	national sample from EEOS	verbal	national sample from EEOS	(-) S*	(-) S*	16,000 + pupils	12	United States
		non-verbal		(-) S*	(-) S*			
		reading		(-) S*	(-) S*			
		math		(-) S*	(-) S*			
		general information		(-) S*	(-) S*			
quadratic expression of teacher's average years experience	national sample from EEOS	verbal	national sample from EEOS	(+) S*	(+) S*	16,000 + pupils	12	United States
		non-verbal		(*) S*	(*) S*			
		reading		(+) S*	(+) S*			
		math		(+) S*	(+) S*			
		general information		(+) S*	(+) S*			

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 8 (continued)

Input Variables		Output Variables		Direction of Effect Reduced Form	Direction of Effect Structural Form	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation					
number of teachers per pupil	national sample from EEOS	verbal	national sample from EEOS	(+) NS	(+) S*	16,000 + pupils	12	United States
		non-verbal		(+) S*	(+) S*			
		reading		(+) S*	(+) S*			
		math		(+) NS	(+) NS			
		general information		(+) S*	(+) S*			
age of school	national sample from EEOS	verbal	national sample from EEOS	(-) NS		16,000 + pupils	12	United States
		non-verbal		(-) S*	(-) S*			
		reading		(+) NS	(+) S*			
		math		(-) NS				
		general information		(-) NS				

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 8 (continued)

Input Variables		Output Variables		Direction of Effect Reduced Form	Direction of Effect Structural Form	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation					
teacher verbal ability score	national sample from EEOS	verbal	national sample from EEOS	(+) S*	(+) S*	16,000 + pupils	12	United States
		non-verbal		(*) S*	(*) S*			
		reading		(+) S*	(+) S*			
		math		(+) S*	(+) S*			
		general information		(+) S*	(+) S*			

Comments:
The use of EEOS (Coleman Report) data from the national sample involved twelfth grade pupils from all regions of the country and different ethnic backgrounds. The Educational Testing Service (ETS) selected the verbal test from the School and College Ability Tests (SCAT), the non-verbal test from the Interamerican Tests of General Ability, the reading and mathematics tests from the Sequential Tests for Educational Progress (STEP), and the general information test from items used in previous ETS research projects. The Coleman researchers identified some tests as measures of ability and some as measures of achievement. Boardman, et al. identified all tests as achievement measures. They concluded that all tests appeared to measure a common characteristic.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

a significant negative effect on non-verbal performance, but also displayed a significant positive effect on reading achievement.

Reciprocal effects between achievement and pupil efficacy and expectations, pupil efficacy and teacher expectations, parent expectations and motivation were theorized. Feedback from achievement to motivation to pupil expectations and back to achievement were theorized. Pupil efficacy and expectations demonstrated a direct effect on achievement, while the other endogenous variables had indirect effects.

The authors concluded that both the home and school are important for all achievement, especially verbal and general information. "Quite substantially these results show that good teachers and good schools are important for educational achievement" (p. 65). Contrary to the Coleman Report, they found that the general information equation best fit the data, not the verbal equation. In fact, they took exception with the general tendency among researchers to use a verbal measure almost exclusively as the measure of achievement.

The use of simultaneous equations with a theory of endogenous relationships and an examination of quadratic effects made this study a uniquely sophisticated review of the Coleman data. Unfortunately, school expenditures were only indirectly examined in these equations although their indirect effects were significant. Also, the analysis remains a captive of the Coleman data, which is substantial but not collected with a production function theory to guide its development. Perhaps the most serious deficiency in this instance is the inability to match pupils and teachers (Michelson, 1970).

Nevertheless, because few studies can match this one in the evolution of the production function model, it remains a standard for researchers to examine.

Sullivan, see Table 9, used correlation analysis to evaluate 614 K-12 school districts in New York state to determine the relationship between the size of the supervisory staff and student achievement in elementary and secondary schools. The author used achievement and socioeconomic data collected by the New York State Education Department in 1971-72 for grades 3, 6, and 10. He employed median family income and per pupil expenditures to control for community and school factors in the analysis.

Sullivan found neither elementary nor secondary supervisory/teacher ratios to have statistically significant correlations with achievement measures in mathematics, reading, or English. He did, however, find significant and positive correlations with achievement between both median family income and per pupil expenditure. Elementary staffing patterns were not related to the socioeconomic nature of the school districts although the secondary patterns were significantly related.

This study used a correlation analysis and, therefore, did not meet the requirements imposed by the production function conceptual model. At the same time, the study attempted to incorporate two widely-used production function proxies to represent the socioeconomic characteristics of the community and school environments. Because model specification has been an important issue in education studies, it is valid to challenge the use of only one proxy for each of two significant variables. The lack of data pertaining to students' family backgrounds seriously impaired the ability of the

Table 9. Study Characteristics of Sullivan

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
supervisor/ teacher ratio in elementary schools	district	reading math	district	(-) NS (-) NS	614 districts	3, 6 3, 6	New York
supervisor/ teacher ratio in secondary schools	district	English math	district	(-) NS (-) NS	614 districts	10 10	New York
median family income	district	reading English math	district	(+) S*** (+) S*** (+) S*** (+) S*	614 districts	3, 6 10 3, 6 10	New York

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 9 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per pupil expenditure	district	reading English	district	(+) S*** (+) S***	614 districts	3, 6 10	New York
		math		(+) S*** (-) S*		3, 6 10	

Comments:
The author found significance to the .001 level=S*** (as shown in table)
The table information is based on correlation data.
Data were collected for the 1971-72 school year at grades 3, 6, and 10.
New York City school districts were excluded from the study because of significant sociodemographic and staff differences with other districts in the state. Eliminating non K-12 districts and those from which a portion of the necessary data were missing reduced the sample size from 756 to 614.
The 1971-72 achievement data for grades 3 and 6 are mean scores on the Pupil Evaluation Program tests. Mathematics and English scores for tenth grade are based on the ratio of pupils passing to number of pupils eligible to take the Regents examinations in Tenth Year Mathematics and English Comprehension.

$*p < .05$ S = SIGNIFICANT
$**p < .01$ NS = NONSIGNIFICANT

study to control this crucial variable. Sullivan suggested that his main thesis may have been undermined by the lack of variability in supervisory/teacher staffing in the districts of New York. Despite these limitations, the results showed that money had a positive and significant effect on achievement.

Summers and Wolfe, Table 10, collected data for three grade levels--6, 8, and 12--for analysis from the Philadelphia school system from 1971-72 and relevant 1970 census information for analyzing block income. They found that pupil background was the dominant contributor to achievement, but that schools would have a more powerful effect if inputs were applied differentially, such as small classes for low achievers and larger classes for more able students. Manipulable school inputs that had a statistically significant impact on achievement included class size, school size, teacher experience, and the quality of the college from which the teacher graduated. The authors suggested that restructuring school budget decisions to reflect these findings would result in educational improvements.

The authors concluded that generally the lack of student-specific data has inhibited production function analyses from demonstrating the effectiveness of school resources more fully. They suggested this to be true because the "interaction" results between income and/or race with school inputs from the rich data set they had on individual students and schools revealed no residual impact of race or income on achievement (except for race in the 12th grade sample). Pedhazur (1982, p. 260), however, quoted Cohen to the effect that the high correlation between the original vectors and their cross

Table 10. Study Characteristics of Summers & Wolf (1974)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
class size	school	ITBS composite	pupils	(-) S* (-) S*	103 42	6 8	Philadelphia
school size	school	ITBS composite	pupils	(-) S* (-) S* (-) S*	103 42 5	6 8 12	Philadelphia
teacher experience	school	ITBS composite	pupils	(+) S* (+) S* (+) S*	103 42 5	6 8 12	Philadelphia
teacher B.A. college ratings	school	ITBS composite	pupils	(+) S* (+) S* (+) S*	103 42 5	6 8 12	Philadelphia
interaction of class size and 3rd grade ITBS score	school	ITBS composite	pupils	(+) S*	103	6	Philadelphia

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 10 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
library books per pupil	school	ITBS composite	pupils	(-) S*	103	6	Philadelphia
interaction of teacher experience with 3rd grade ITBS score	school	ITBS composite	pupils	(+) S*	103	6	Philadelphia

Other Comments:
This chart was selected from the findings for regression analyses using a value added approach with reference to Charts D-1, D-2, and D-3 in the study. The significance of the variables was based on t-test results from the regression equations. The sample of pupils included 627 6th graders from 103 elementary schools, 553 8th graders from 42 junior high schools, and 716 12th grade pupils from 5 senior high schools. Because data were aggregated at the school level for school level analyses, it is the number of schools that must be identified as the appropriate sample size.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

products result in "arbitrary nonsense" when included in a simultaneous analysis such as Summers and Wolfe performed.

A significant characteristic of the study was the use of a value-added approach in measuring student achievement. By matching prior standardized test results for individual students with their grade level test scores to control for innate ability, prior schooling inputs, and family background, the contribution of specific school resources could be more accurately assessed according to the conceptual model. Unfortunately, the tests were not in every case identical and different measures for tests were used. For example the 12th grade percentile score on the California Aptitude Test was paired with the 9th grade percentile score on the Cooperative School and College Abilities Test and the 7th grade ITBS grade equivalent score. For the 6th grade, Iowa Test of Basic Skills grade equivalent scores were used. Standard scores would provide the best data for comparative analysis, but they were not selected. Pedhazur pointed out that grade equivalents were not on an equal interval scale and "[t]his in itself renders them of dubious value as a measure of the dependent variable, not to mention the further complication of using the differences between such measures" (1982, p. 259).

Pedhazur further expressed concern over the problem of specification errors, the meaningfulness of college rating scales used in the analysis, and contradictions in the findings between this study and a similar study using the same methodology in 1977. Ultimately, however, "[t]he most important criticism of the Philadelphia studies is that they are devoid of theory..." (p. 259). The close relationship to the conceptual model based on the variables used in the regression equations, could not overcome the deficit of theory. Moreover, Pedhazur considered the use of such studies for policy decisions to be dangerous. "[T]he unintended damage of conclusions and actions based on questionable research designs and the inappropriate use of analytic methods is incalculable" (p. 264).

Nichols, Table 11, used information accumulated by the New York State Education Department for the Basic Education Data System during the 1972 academic year. In addition to information collected from each district, data from the 1970 United States Census aggregated for the 734 districts in the state were applied to each school in each district. Census data were aggregated separately for 31 sub-districts in New York city. Proxies were used to measure characteristics of school educational programs, teacher characteristics aggregated by school, district level financial resources, characteristics of the school pupil population. Dummy variables were entered into the equation to identify school location. All proxies used in the study were not included in the chart.

The author divided proxies on the input side of the equation into input and treatment variables. Inputs represented pupil characteristics "primarily the economic, education and racial-ethnic status of the students' homes" (p. 8). Treatment variables included per pupil expenditures and teacher/school characteristics. Per pupil expenditures along with teacher experience, age, and salary were statistically significant correlations with the mean Pupil Evaluation Program (PEP) achievement scores. Since treatment and input variables as identified by the author are frequently correlated, partial correlations were reported in order to control for student background variables. In this case, the per

Table 11. Study Characteristics of Nichols

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
expenditure per pupil	district	composite PEP score	school	(+) S** (+) S**	1701 1433	3 6	New York
pupil-teacher ratio	school	composite PEP score	school	(+) S** (+) S**	1701 1433	3 6	New York
% children from families on welfare	school	composite PEP score	school	(-) S** (-) S**	1701 1433	3 6	New York
school location in New York city	school	composite PEP score	school	(-) S** (-) S**	1701 1433	3 6	New York
school location in suburb of large city	school	composite PEP score	school	(+) S** (+) S**	1701 1433	3 6	New York
mean years of teacher experience	school	composite PEP score	school	(+) S** (+) S**	1701 1433	3 6	New York

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 11 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
mean years of teacher experience in district	school	composite PEP score	school	(+) S** (+) S**	1701 1433	3 6	New York
mean family income	district	composite PEP score	school	(+) S** (+) S**	1701 1433	3 6	New York
% male and female with 4 yrs. high school	district	composite PEP score	school	(+) S** (+) S**	1701 1433	3 4	New York

Other Comments:
Mean composite scores from the New York state Pupil Evaluation Program (PEP) that include reading and mathematics results from 1972 were used in the analysis. The findings reported in this chart are based on correlations between school characteristics and mean PEP scores. School level data were used for most of the variables. Census (1970) data were used for determining district level characteristics for mean family income, adult educational attainment, and per pupil expenditures.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

pupil expenditures were not significant, but the teacher characteristics remained significant at the .05 level.

Economic and demographic characteristics of the pupils contributed the most (R^2=.70) to an explanation of the variance in test performance between schools. All input variables contributed .84 to third grade performance and .85 for sixth grade performance. Nichols, however, used stepwise regression as an explanatory tool in this analysis. This was inappropriate since stepwise regression is solely a predictive tool (Pedhazur, 1984). The study met the criteria for a traditional empirical production function study within the circumscribed availability of appropriate data. The study was cross-sectional and lacked any measure of previous pupil learning or ability. Consequently, the input controls were exclusively demographic and not academic. In addition, the high correlations between proxies that were acknowledged by the author, undermined the value of conclusions even when partial correlations were used. Although the study was performed with skill and it could be regarded as one of the most theoretically complete in this dissertation, it still could not measure up to the conceptual model described in Chapter II.

Murnane and Phillips, Table 12, investigated two instructional methodologies as well as teacher effectiveness with data collected as part of a federally-funded welfare reform experiment in inner-city schools. All of the families were African-American, more than one-half were headed by females, and most had very low incomes. The authors used multivariate procedures from the production function and process-product approaches. The findings from both methodologies identified effective teacher indicators.

Table 12. Study Characteristics of Murnane and Phillips

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher experience = 7 years or less	classroom	vocabulary	classroom	(+) NS	171 pupils 23 teachers	3	large midwestern city
				(+) S*	176 pupils 20 teachers	4	
				(-) NS	159 pupils 22 teachers	5	
				(+) S*	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher experience = 8 to 14 years	classroom	vocabulary	classroom	(-) S*	171 pupils 23 teachers	3	large midwestern city
				(-) NS	176 pupils 20 teachers	4	
				(+) NS	159 pupils 22 teachers	5	
				(-) NS	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher experience = 15 years or more	classroom	vocabulary	classroom	(+) S*	171 pupils 23 teachers	3	large midwestern city
				(+) S*	176 pupils 20 teachers	4	
				(+) NS	159 pupils 22 teachers	5	
				(+) S*	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teachers with a master's degree	classroom	vocabulary	classroom	(+) NS	171 pupils 23 teachers	3	large midwestern city
				(+) NS	176 pupils 20 teachers	4	
				(+) NS	159 pupils 22 teachers	5	
				(-) NS	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher verbal ability score	classroom	vocabulary	classroom	(+) NS	171 pupils 23 teachers	3	large midwestern city
				(-) NS	176 pupils 20 teachers	4	
				(-) NS	159 pupils 22 teachers	5	
				(-) S*	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher attended prestigious college	classroom	vocabulary	classroom	(+) NS	171 pupils 23 teachers	3	large midwestern city
				(+) NS	176 pupils 20 teachers	4	
				(+) NS	159 pupils 22 teachers	5	
				(+) NS	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher is white	classroom	vocabulary	classroom	(+) NS	171 pupils 23 teachers	3	large midwestern city
				(-) NS	176 pupils 20 teachers	4	
				(-) NS	159 pupils 22 teachers	5	
				Insufficient data	168 pupils 17 teachers	6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 12 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher is male	classroom	vocabulary	classroom	Insufficient data	171 pupils 23 teachers	3	large midwestern city
				(-) NS	176 pupils 20 teachers	4	
				(+) S*	159 pupils 22 teachers	5	
				(-) S*	168 pupils 17 teachers	6	

Other Comments:
The findings were based on regression coefficients. All pupils were African-American.
The vocabulary component of the Iowa Test of Basic Skills for the appropriate grade level was the achievement measure. The classrooms were self-contained at the elementary school level.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Coding dummy variables for classrooms allowed the authors to match pupils with their teachers.

Using six available measures of teacher characteristics (see Table 12), the authors found that teaching experience was related to pupil achievement systematically, but in a nonlinear manner. Teacher experience was subdivided using a "three piece linear spline specification" (p. 29). Historical labor conditions in education explained the findings on experience. Teachers with experience in the range from 8-14 years were hired during a period of rapid pupil population growth. Those hired before and after rapid pupil population growth were more selectively chosen because the pressure to hire was much lower. These circumstances suggest that the innate abilities of teachers employed may vary as a consequence of the magnitude of personnel needs of school districts. Because teacher verbal ability was determined by a self-administered test and teachers used aids, the authors viewed the results as spurious. The small number of male teachers contributed to inconsistent results in the analysis. The other characteristics did not demonstrate any relationship to pupil achievement.

The authors found the direct instruction approach to be especially effective with inner-city pupils, but "successful implementation of this approach requires that the teacher find the techniques that provide the best match of his or her skills and personality with the curriculum, the available materials, and most importantly, with the needs of the children in the class" (p. 35). Experienced teachers and more able teachers are most effective in making these decisions.

The authors employed a linear regression model similar to the conceptual model identified in Chapter II. A vector of teacher characteristics was separated from general school characteristics, and achievement at the conclusion of the prior school year was entered into the equation in order to control for innate abilities and previous achievement. While this empirical analysis closely follows the conceptual model, the authors regard the use of the term "education production function" unfortunate because the term "has a very specialized meaning to economists which is inappropriate for studying the relationship between teachers and pupil achievement" (p. 2).

Comparing metropolitan and nonmetropolitan school districts in Alabama, statistically significant achievement differences related to the percentage of African-American pupil enrollment and the type of school system (city or county). The results were negatively related to African-American pupil enrollment and city systems produced higher achievement. In nonmetropolitan school districts, per pupil expenditures for eighth grade pupils and average daily attendance for twelfth grade pupils were statistically significant and positive. In metropolitan school districts, per pupil expenditures and teacher-pupil ratio had a statistically significant and positive effect on eighth grade achievement. The analysis controlled for socioeconomic differences in order to concentrate on structural characteristics of schools.

Dunkelberger and Soderberg, Table 13, proposed a causal model following the work of Bidwell and Kasarda (1975) that used the school system as the unit of analysis with metropolitan and nonmetropolitan service areas as the key environmental variable. Seven independent variables were used in the model: expenditure per pupil; system size;

Table 13. Study Characteristics of Dunkelberger and Soderberg

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
expenditure per pupil	district groups--metro & nonmetro	composite score	district groups--metro & nonmetro	(+) S* (+) S** (+) NS (+) NS	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama
pupil-teacher ratio	district groups--metro & nonmetro	composite score	district groups--metro & nonmetro	(+) NS (-) S* (+) NS (+) NS	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama
average daily attendance	district groups--metro & nonmetro	composite score	district groups--metro & nonmetro	(+) NS (+) NS (+) S* (-) NS	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama
average high school teachers' salary	district groups--metro & nonmetro	composite score	district groups--metro & nonmetro	(+) NS (+) NS (+) NS (+) NS	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 13 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
system size (enrollment)	district groups--metro & nonmetro	composite score	district groups--metro & nonmetro	(+) NS (+) NS (+) NS (+) NS	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama
system type	district groups--metro & nonmetro	composite score	district groups--metro & nonmetro	(+) NS (+) NS (+) NS (+) NS	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama
% of African-American pupils	district groups--metro & nonmetro	composite	district groups--metro & nonmetro	(-) S** (-) S** (-) S** (-) S**	45 (nonmetro) 82 (metro) 45 (nonmetro) 82 (metro)	8 8 12 12	Alabama

Other Comments:
Mean composite scores that include reading, spelling, and mathematics for the 1979 California Achievement Test results were used in the analysis. The findings reported in this chart are based on regression coefficients. Districts were grouped into metropolitan and nonmetropolitan systems on the basis of Standard Metropolitan Statistical Areas.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

average daily attendance; pupil-teacher ratio; system type; average high school teacher salary; and percent of African-American pupils. Statistical analyses involved the use of t-tests to identify the extent of mean variation test scores between metropolitan and nonmetropolitan school systems, correlations to examine the relationships between the seven independent variables and mean test scores, and linear multiple regression to estimate the effects of each structural variable on test scores. The authors reported an R^2 of .81 for eighth graders and .41 for nonmetropolitan schools (significant to .001). The seven independent variables explained .77 of the variance in eighth grade achievement and .71 of the twelfth grade achievement (significant to .01) in metropolitan districts.

Although the researchers proposed a model for analysis, it did not encompass the variables of a full conceptual design. No estimates of pupil ability were included. Peer background was limited to percent of African-American pupils. The researchers appeared to use the percent African-Americans as the exclusive SES measure, and the average high school teacher salary as the exclusive measure of teacher quality. It is most important to understand that the model R^2 explained the variance between metropolitan and nonmetropolitan groups having very small sample sizes--the nonmetrpolitan group was 45 and the metropolitan group was 82. The level of data aggregation permits no further explanation of the effect of the variables. It is instructive, however, that per pupil expenditures and class size show significance at this level of analysis. Although not conceptually complete, the model supports the importance of educational expenditures in pupil achievement.

Wendling and Cohen, Table 14, found operating expenditures per pupil and instructional expenditures per pupil to have a positive and statistically significant relationship to pupil achievement. Teacher quality as identified by experience and master's degree level showed a similar result. Median years of schooling in district populations also exhibited a significant and positive relationship with achievement. While school size was not a statistically significant factor, the percent of pupils below poverty level and minority status displayed negative and statistically significant relationships to achievement. These relationships were consistent for both reading and mathematics performance.

Among the geographical groupings that included rural schools, suburban schools, upstate city schools, and New York city schools, rural schools displayed certain unique and unexplained differences. Median years of schooling had no effect on pupil achievement, and the percent of the population below the poverty level, although 2% higher than the state average, showed little effect on mathematics performance. The average achievement levels were slightly higher than the state sample, but teacher experience was one year less, real expenditures were 9% lower, and instructional expenditures were 6% lower.

The authors identified high and low achieving schools based on a one standard deviation above and below the state mean on third grade achievement test performance. For low achieving schools, approved expenditures were 15% less and instructional expenditures were 25% less per pupil than high achieving schools. The percent of minority pupils was 17 times that of high achieving schools. Ninety percent of the low achieving schools were in central cities and 87% of the high achieving schools were located in suburban communities.

The authors concluded that schools make a difference for pupil learning and that distributing additional funds to low achieving schools should lead to an increase in academic performance. Rural schools should be investigated further to attempt to explain their better than expected performance. Because home factors are important to pupil achievement, schools need to identify demographic characteristics of their communities and promote ways to work with parents in helping pupils succeed in school. Also, factors that are effective in working with minority children should be an area of focus for schools.

Wendling and Cohen did not state a conceptual model, but they selected proxies associated with variables described in Chapter II, except for a measure of pupil ability. The data used in the study were cross-sectional. The authors edited the data to include schools as homogenous as possible except for the input resources evaluated. The variance in geographical effects resulting from a variety of demographic factors should caution researchers to avoid confounding analyses through data selection.

Vanvalkenberg, Table 15, found statistically significant relationships between student performance on the Michigan Educational Assessment Program (MEAP) and several financial variables. Test scores for fourth, seventh, and tenth grade students in reading and mathematics were evaluated from 1984 and 1985. Science scores from 1986 were examined. The sample was confined, however, to the Wayne County area that included 36 districts in 1984 and 34 districts in 1985 and 1986. Community wealth,

Table 14. Study Characteristics of Wendling and Cohen

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher average experience	school	reading math	school	(+) S* (+) S*	1,021 schools	3 3	New York
teacher % Master's degree	school	reading math	school	(+) NS (+) S*	1,021 schools	3 3	New York
% minority	school	reading math	school	(-) S* (-) S*	1,021 schools	3 3	New York
school size	school	reading math	school	(-) NS (-) NS	1,021 schools	3 3	New York
approved operating expenditures per pupil	district	reading math	school	(+) S* (+) S*	1,021 schools	3 3	New York

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 14 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
instructional expenditures per pupil	district	reading math	school	(+) S* (+) S*	1,021 schools	3 3	New York
median years of schooling	district	reading math	school	(+) S* (+) S*	1,021 schools	3 3	New York
% below poverty level	district	reading math	school	(-) S* (-) S*	1,021 schools	3 3	New York

Other Comments:
Data were taken from the Consolidated Data Base of the New York State Eduction Department's Basic Educational Data System (BEDS) for the sample of 1,021 schools. Achievement data (dependent variables) came from the 1977-78 school year. Standardized test scores for third grade reading and mathematics composites were averaged for each school in the sample. Schools were also aggregated by four regional groupings for analysis. Regression coefficients and t-scores were reported for each independent variable.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 15. Study Characteristics of Vanvalkenburg

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
adjusted gross income (1984)	district	reading	district	(+) S*	36	4,7,10	Wayne County, Michigan
		math		(+) S*		4,7,10	
adjusted gross income (1985)	district	reading	district	(+) S*	34	4,7,10	Wayne County, Michigan
		math		(+) S*		4,7,10	
residential equalized assessed valuations valuation (1984)	district	reading	district	(+) S*	36	4,7,10	Wayne County, Michigan
		math		(+) S*		4,7,10	
residential equalized assessed valuation (1985)	district	reading	district	(+) S*	34	4,7,10	Wayne County, Michigan
		math		(+) S*		4,7,10	
residential equalized assessed valuation (1986)	district	science	district	(+) S*	34	4,7,10	Wayne County, Michigan

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 15 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
instructional expenditures per pupil (1984)	district	reading	district	(+) S* (+) NS	36	7,10 4	Wayne County, Michigan
		math		(+) S* (+) NS		7,10 4	
instructional expenditures per pupil (1985)	district	reading	district	(+) S* (+) NS	34	4,10 7	Wayne County, Michigan
		math		(+) S* (+) NS		7,10 4	
instructional expenditures per pupil (1986)	district	science	district	(+) NS	34	4,7,10	Wayne County, Michigan
size of district (1984)	district	reading	district	(-) NS	36	4,7,10	Wayne County, Michigan
		math		(-) NS (+) NS		4,7 10	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 15 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
size of district (1985)	district	reading	district	(-) NS (+) NS	34	4,7 10	Wayne County, Michigan
		math		(-) NS (+) NS		4,7 10	
size of district (1986)	district	science	district	(+) NS (-) NS	34	4,10 7	Wayne County, Michigan
minority enrollment (1984)	district	reading	district	(-) S*	36	4,7,10	Wayne County, Michigan
		math		(-) S*		4,7,10	
minority enrollment (1985)	district	reading	district	(-) S*	34	4,7,10	Wayne County, Michigan
		math		(-) S*		4,7,10	
minority enrollment (1986)	district	science	district	(-) S*	34	4,7,10	Wayne County, Michigan

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 15 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
pupil/teacher ratios (1984)	district	reading math	district	(-) S* (-) S*	36	average of 4,7,10	Wayne County, Michigan
pupil/teacher ratios (1985)	district	reading math	district	(-) S* (-) S*	34	average of 4,7,10	Wayne County, Michigan
pupil/teacher ratios (1986)	district	science	district	(-) NS	34	average of 4,7,10	Wayne County, Michigan

Other Comments:
The Michigan Educational Assessment Program (MEAP) evaluates basic student competencies in reading, mathematics, science, and health. The MEAP criterion-referenced tests are designed to provide information on student skills that can be utilized in the instructional process. Music ability and other skills have been assessed with small district samples during the operation of the program.
Correlations were reported for this analysis.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

represented by adjusted gross income (AGI) and state equalized valuation (SEV) was positive and significant for all grades and subjects. These findings tend to reinforce the view that community wealth correlates with student performance. Instructional expenditures per pupil were positive and significant for reading and mathematics in grades seven and ten for 1984 and 1985, but not significant for grade four both years. The relationship was not significant for science in 1986. Minority enrollment showed a significant negative correlation across all subjects and grades. Pupil-teacher ratios were significantly negative across all grades and subjects except science. District size was not significant for all correlations.

While the study deviates from the regression model, it did examine data encompassing community and demographic variables. The level of analysis, however, remained at the district level. Consequently, the sample size, although encompassing a large student population, was limited to the number of districts in Wayne County and was quite small (34-36 districts). The MEAP, like criterion-referenced tests from other states examined in this study, did demonstrate a significant relationship with the variables used in the analysis.

Hokanson, Table 16, used path analysis to evaluate the relationship between school expenditures and pupil achievement for third graders in Florida's public schools. He hypothesized an explicit causal relationship using financial and demographic data from the Florida Department of Education. Proxies for socioeconomic (SES) variables were taken from the annual Civil Rights Survey and principals' reports to the Department of Education; funding proxies were reported through the Florida Education Finance

Table 16. Study Characteristics of Hokanson

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent pupils with white collar head of household	school	combined math and communication scores	school	(+) S*	1125	3	Florida
percent pupils eligible for free and reduced lunch	school	combined math and communication scores	school	(-) S*	1125	3	Florida
percent African-American pupils	school	combined math and communication scores	school	(-) S*	1125	3	Florida
teacher salaries and benefits	school	combined math and communication scores	school	(+) S*	1125	3	Florida

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 16 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
professional and technical services by non-School Board employees	school	combined math and communication scores	school	(+) S*	1125	3	Florida
expenses for dues, fees, and non-contractual substitute teachers	school	combined math and communication scores	school	(-) S*	1125	3	Florida
capital outlay for audio-visual equipment and materials	school	combined math and communication scores	school	(+) S*	1125	3	Florida
instructional support services: administration, guidance, and health	school	combined math and communication scores	school	(-) S*	1125	3	Florida

$^{*}p < .05$ S = SIGNIFICANT
$^{**}p < .01$ NS = NONSIGNIFICANT

Table 16 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
Title 1 salaries and benefits	school	combined math and communication scores	school	(-) S*	1125	3	Florida
Title 1 expenses	school	combined math and communication scores	school	(+) S*	1125	3	Florida

Comments:
The independent variables in this chart are from the final (reduced) regression model for a path analysis.
The data sample included all third grades in the public schools of Florida that had a second grade for which instructional funds were expended in the 1979-1980 school year (1125).
The test measure was a school's average score in mathematics and communication combined for all third graders on the September, 1980 Statewide Assessment Test, a criterion-referenced test developed for Florida.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Program (FEFP); and achievement test scores were taken from the 1980 Florida State Assessment results. Financial proxies were classified as state and locally funded or federally funded (Title 1). In a careful exposition on the tenuous nature of causal analysis in a naturalistic setting, Hokanson found justification for his analytical approach in the professional literature.

The author found a positive and significant relationship with achievement for the salaries and benefits paid to teachers, the services provided by professional and technical non-School Board employees, capital outlay for audio-visual equipment and materials, and Title I (now Chapter 1) instructional expenses (excluding salaries). Hokanson found a significant negative relationship for expenditures for dues, fees, and non-contractual substitute teachers, instructional support services such as administration, guidance, and health personnel, and Title 1 salaries and benefits. Several demographic factors were significant: the percent of students whose head of household was a white collar worker was positive; the percent of pupils eligible for free and reduced lunch and the percent of pupils who were African-American were both negative.

Although Hokanson used path analysis to theorize causal relationships, multiple linear regression was the primary tool for testing relationships among the variable proxies and an education production function model was used. Comparing his model with the conceptual model, no proxy for student ability was used. SES characteristics cannot satisfy the deficiency. The source of information for the significant positive SES characteristic, percent of pupils whose parents are white collar workers, comes from estimates by school principals. Consequently, the reliability of such data can be challenged. Hokanson expressed concern about the ceiling effect created by testing minimum skill levels that truncated achievement scores at higher levels and the use of a school mean as the unit of measurement. The combination of mathematics and communication scores provided a unique measure of pupil achievement.

The purpose of Brock's study, Table 17, was to examine selected socioeconomic (SES) and per pupil expenditure variables and their relationship to pupil achievement in reading and mathematics as measured by CTBS-U standardized tests. Data were evaluated for 76 unitary school districts in Kentucky for the 1982-83 school year. Brock performed a correlation analysis of the two selected SES variables, personal income and level of education, with pupil scores. He then partialed out the SES factors in a multiple regression analysis to determine which expenditure variables contributed to student achievement over and above the influence of SES.

Brock found instructional expenditures to have a positive and significant effect on student achievement in seven of eight grade/subject analyses when controlling for SES factors. Total expenditures, salaries for teachers, support personnel, and principals were also positive and significant in one or more analyses. The author found a high correlation between the SES factors, personal income and level of education, and student achievement on CTBS-U mathematics and reading tests at grades 3, 5, 7, and 10. The education of parents appeared to have a higher correlation with achievement than income. Achievement test scores were higher in mathematics in districts with higher teacher salaries, although this was not necessarily the case for reading scores. The gap between student achievement scores for schools that spent more on per pupil instructional costs

Table 17. Study Characteristics of Brock

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent parents with a high school diploma	district	reading	district	(+) S**	76 districts	3, 5, 7, 10	Kentucky
		math		(+) S**		3, 5, 7, 10	
per capita personal income	district	reading	district	(+) S**	76 districts	3, 5, 7, 10	Kentucky
		math		(+) S**		3, 5, 7, 10	
LFI (index of local effort to support education)	district	reading	district	(+) NS (-) NS	76 districts	3, 7, 10 5	Kentucky
		math		(+) NS		3, 5, 7, 10	
superintendent's cost	district	reading	district	(+) NS	76 districts	3, 5, 7, 10	Kentucky
		math		(+) NS (-) NS		3, 7, 10 5	

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 17 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
salaries of principals and assistant principals	district	reading	district	(-) NS (+) NS	76 districts	3 5, 7, 10	Kentucky
		math		(+) NS (+) S**		3, 7, 10 5	
teacher salaries	district	reading	district	(+) NS (-) NS	76 districts	3, 5, 7 10	Kentucky
		math		(+) S** (+) S# (+) NS		7 3, 5 10	
salaries of support personnel in guidance and libraries	district	reading	district	(+) NS	76 districts	3, 5, 7, 10	Kentucky
		math		(+) NS (+) S#		3, 10 5, 7	
cost of library books, materials, audiovisuals, teaching supplies	district	reading	district	(+) NS (-) NS	76 districts	3, 10 5, 7	Kentucky
		math		(-) NS (+) NS		3, 5, 7 10	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 17 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
total instructional expenditures	district	reading	district	(+) S# (+) S* (+) NS	76 districts	3, 5 7 10	Kentucky
		math		(+) S* (+) S** (+) S#		3, 5 7 10	
total expenditures	district	reading	district	(+) S* (+) NS	76 districts	3, 7 5, 10	Kentucky
		math		(+) S** (+) S* (+) NS		3 7 5, 10	

Comments:
The chart data for expenditure variables reflect the partialing out of the socioeconomic proxies (income and education) in the regression analysis. The National Curve Equivalent (NCE) scores from the mathematics and reading sections on the CTBS-U standardized test are the dependent variables. Data were collected for the 1982-83 year at grades 3, 5, 7, and 10. The level of significance was set at .10, represented in the table as S#.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

compared to those who spent less grew each year during the elementary years. This gap did not appear to widen at the secondary level.

Brock used data from the Kentucky Department of Education and the 1980 census. Since school districts and counties were identical units, census data provided valid SES information at the level of aggregation analyzed. Some changes could be expected to occur, however, in the two year gap between census and district data. While the author cannot be faulted for using the reliable data available, one must question if these two mean values adequately capture the SES of the school district. Does the school population match the overall demographic characteristics of the county? Do private schools pull out a significant number of higher SES pupils? Is there a higher number of students from the lower income and education population group in the public schools than is represented in the county as a whole? The percent of free and reduced lunch students could provide at least one useable peer group characteristic for the study. Comparing the proxies selected for this study with the conceptual model, the missing variables for peer group and pupil ability level are notable. One can question which categories, family background or peer characteristics, the SES proxies may properly identify in relationship to the conceptual model given the potential discrepancies between county and school data. Finally, the analysis is subject to the partialing fallacy where the findings may encompass several explanatory models (Pedhazur, 1982). Ultimately, this study falls prey to the availability of data, fulfillment of the conceptual design, and the level of aggregation that continually plague all education production function analyses.

Lavalley, Table 18, found significant relationships between several finance variables and pupil achievement in Oklahoma. Instructional expenditures, district level fixed charges, transportation costs, and district level wealth as measured by per capita valuation demonstrated positive and significant relationships with the Metropolitan Achievement Test (MAT). Library expenditures and administrative expenditures demonstrated a significant negative relationship to pupil achievement. District size displayed a non-significant relationship to achievement.

The author used a step-wise multiple regression methodology to investigate input-output relationships. The study was designed to determine how Oklahoma school districts allocated resources and how they were related to student achievement. F-values, correlations, and model R^2 values were reported. Unfortunately, step-wise procedures are useful only for prediction, not for explanation (Pedhazur, 1984). The author, however, did propose the development of a predictive model for pupil achievement, but the R^2 values were too small to generate a useful design.

Lopus, Table 19, tested the effect of expenditures at three levels of aggregation on student achievement in high school economics classes using a traditional production function approach. She found total expenditures, instructional expenditures, and disaggregated expenditure measures to be statistically significant. Thus, expenditures at every level of aggregation were significant. Indeed, she found that the more disaggregated the measure or the more directly targeted to the instruction of the specific students evaluated, the more influence expenditures have on achievement.

Table 18. Study Characteristics of Lavalley

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
instruction	district	MAT	district	(+) S** (+) S** (+) S**	611 districts	3 7 10	Oklahoma
fixed charges	district	MAT	district	(+) S* (+) S** (+) S**	611 districts	3 7 10	Oklahoma
library	district	MAT	district	(-) S**	611 districts	10	Oklahoma
administration	district	MAT	district	(-) S** (-) S** (-) S**	611 districts	3 7 10	Oklahoma
transportation	district	MAT	district	(+) S** (+) S** (+) S**	611 districts	3 7 10	Oklahoma
per capita valuation	district	MAT	district	(+) S** (+) S** (+) S**	611 districts	3 7 10	Oklahoma

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 18 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
school district size	district	MAT	district	(+) NS (+) NS (+) NS	611 districts	3 7 10	Oklahoma

Other Comments:
The Metropolitan Achievement Test, Sixth Edition, was used to test 110,000 Oklahoma pupils in grades three, seven, and ten during 1986. The use of the norm-referenced test initiated the first state-wide testing program for Oklahoma. Correlation data are reported in the chart. Instruction included activities of the teacher, principal, consultant, supervisor of instruction, guidance personnel, and psychologists. Fixed charges included retirement funds, social security, pension payments, tax-sheltered annuities, insurance, and judgments, rental of land or buildings, interest on warrants, and refunds or penalties--items that do not fit easily into any other category. Library expenditures were available only for high schools.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 19. Study Characteristics of Lopus

Input Variables		Output Variables		Direction of Effect (1,2,3 = equation)	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
total expenditure per pupil	state	Test of Economic Literacy	class	1 (+) S**	2440	9-12	US
instructional expenditure per pupil	state/ district	Test of Economic Literacy	class	2 (+) S**	2440	9-12	US
economics teachers with advanced degrees	class	Test of Economic Literacy	class	3 (+) S**	2440	9-12	US
economics teacher experience	class	Test of Economic Literacy	class	3 (-) NS	2440	9-12	US
economics class size	class	Test of Economic Literacy	class	3 (-) S**	2440	9-12	US
use of audiovisuals	class	Test of Economic Literacy	class	3 (+) NS	2440	9-12	US

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 19 (continued)

Input Variables		Output Variables		Direction of Effect (1,2,3 = equation)	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
use of computers	class	Test of Economic Literacy	class	3 (-) NS	2440	9-12	US
use of supplementary texts		Test of Economic Literacy	class	3 (-) S*	2440	9-12	US
use of consultants or services	class	Test of Economic Literacy	class	3 (+) S**	2440	9-12	US
sex of pupil	class	Test of Economic Literacy	class	1 (+) S* 2 (+) S** 3 (+) S*	2440	9-12	US
race of pupil	class	Test of Economic Literacy	class	1 (+) S** 2 (+) NS 3 (+) S**	2440	9-12	US
pupil grade point average	class	Test of Economic Literacy	class	1 (+) S** 2 (+) S** 3 (+) S**	2440	9-12	US

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 19 (continued)

Input Variables		Output Variables		Direction of Effect (1,2,3 = equation)	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
pupil in 12th grade	class	Test of Economic Literacy	class	1 (-) S* 2 (-) NS 3 (-) NS	2440	9-12	US
percentage of pupils planning to attend college	class	Test of Economic Literacy	class	1 (+) NS 2 (-) NS 3 (+) S**	2440	9-12	US
teacher attitude	class	Test of Economic Literacy	class	1 (+) S** 2 (+) S** 3 (+) S**	2440	9-12	US

Comments:
The Test of Economic Literacy (TEL) is a nationally-normed, 46-question, multiple choice test designed to measure the economic knowledge and reasoning of eleventh and twelfth grade high school pupils. TEL scores and demographic data were taken from a 1987 administration of the test and accompanying survey instrument. Expenditure data were taken from statewide averages published by the National Center for Educational Statistics for 1987.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Using 1987 data from the National Assessment of Economic Education (NAEE) Survey Database and statistics from the U.S. Department of Education, Lopus defined her empirical model in the following way:

Score on TEL = f(aptitude, socioeconomic factors, peer group effects, teacher attitude).

With the addition of school factors represented by disaggregated expenditures, this model closely approaches the conceptual model discussed in Chapter II with a few exceptions. The aptitude score was based on the sample students' Grade Point Averages (GPA), since no general intelligence rating was available from the data, and a teacher attitude variable was added. Using the GPA actually gives the model a value-added design. The result is a conceptually satisfying structure within the traditional regression methodology. Expenditure data, however, were aggregated at the state level. Using this data in comparisons with school-level data can introduce significant errors of measurement and interpretation.

Coffey, Table 20, found statistically significant correlations between financial expenditures and pupil performance on criterion-referenced tests in mathematics and language arts. She identified positive and significant results for per pupil expenditures for regular instruction, total current costs, and administrative services. Teacher-pupil ratios as well as community demographic characteristics (district wealth, occupational status, and educational background) also demonstrated a positive and significant correlation with pupil scores on the 1991 Connecticut Mastery Test (CMT). Significant negative findings were found for per pupil special education, compensatory education,

Table 20. Study Characteristics of Coffey

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
district wealth	school district	language arts math	school district	(+) S* (+) S*	147	8	Connecticut
per pupil compensatory education	school district	language arts math	school district	(-) S* (-) S*	147	8	Connecticut
per pupil regular instruction	school district	language arts math	school district	(+) S* (+) S*	147	8	Connecticut
per pupil total current expenditures	school district	language arts math	school district	(+) S* (+) S*	147	8	Connecticut
per pupil administrative services	school district	language arts math	school district	(+) S* (+) S*	147	8	Connecticut
per pupil special education	school district	language arts math	school district	(-) S* (-) S*	147	8	Connecticut

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 20 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per pupil support services	school district	language arts math	school district	(+) NS (+) NS	147	8	Connecticut
mean teacher salary	school district	language arts math	school district	(+) NS (+) NS	147	8	Connecticut
equalization aid levels	school district	language arts math	school district	(-) S* (-) S*	147	8	Connecticut
teacher per pupil ratios	school district	language arts math	school district	(+) S* (+) S*	147	8	Connecticut

Comments:
The test scores for language arts and mathematics (math) are taken from the results of the October, 1988, performance of eighth grade pupils on the Connecticut Mastery Test, a criterion-referenced test of achievement. Scores represent the average number of objectives mastered by pupils in each district.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

and pupil support service expenditures. Similarly, the level of state equalization aid demonstrated a negative correlation with test performance.

The author used community demographic characteristics and district expenditures, but other elements of the conceptual model were absent from the study. No measures of parent characteristics were identified. The demographic characteristics used in the study represented adults with children in school and those with none. Likewise, the occupational status measures were for those employed from age 16. The educational attainment measure represented the percent of adults above 25 who had completed college. These data are nonuniform and could have a confounding effect on the correlations with achievement. Analyses at the district level have been questioned for producing aggregation biases. The average number of objectives mastered for each division on the CMT served as the measure of achievement. This unit avoided the truncation of data by pass/fail grading that leads to attenuated variances often observed in state competency testing.

Studies at the district level frequently do not result in positive and significant findings on the relationship between educational expenditures and pupil achievement (Hanushek, 1986; Bridge, Judd, & Moock, 1982). This study, however, presented findings that provide support for the view that certain educational expenditures may have a positive effect on achievement, although the level of state equalization aid did not compensate for a lack of district wealth or low education and occupational status in district communities.

Elizabeth Holtzapple, Table 21, used hierarchical linear modelling for her analysis of student performance in southwest Pennsylvania schools. She selected a sample of 170 schools from the six county greater Pittsburgh area. Holtzapple attempted to merge the conceptual and methodological research on effective schools, student achievement, and education production functions in order to determine how and to what extent schools affect student achievement given a specific SES-achievement relationship. Table 21 presents the variables that fit the final school-level prediction models and explains characteristics of the achievement measures. In her analysis the author developed both individual and school level models. "The school level regression models explain a much larger percentage of the variances in the TELLS outcome measures than do the individual level models. In addition to much larger R squared values, the school level models require fewer independent variables to explain greater amounts of variance in the dependent variables" (p. 136).

Several financial measures were in the final school level models. District wealth--a composite that included measures of local revenues as a percentage of total district revenues, the Market Value Personal Income Aid Ratio, and regular instructional expenditures per weighted average daily membership--demonstrated a positive and significant effect on third grade total and mathematics performance scores. Teacher preparation time--a cost to the district that requires additional personnel--demonstrated a positive and significant effect on student performance in mathematics and total achievement at the eighth grade level. The percent of low income students demonstrated a negative and significant relationship to student achievement in fifth and eighth grade measures of reading, mathematics, and total achievement.

This study has unusual strengths that push the quality of production function studies to a higher qualitative level. Quantitative measures from state administered achievement tests provided individual district, school, and student level data. Questionnaire information from teachers and administrators in the schools under study provided information about characteristics related to teacher involvement in decision making and effective schools characteristics. The use of hierarchical linear modeling (HLM) provided multi-level models in an attempt to deal with the problems resulting from cross-level aggregated data. HLM provides for the formulation of a structural model at each level of the hierarchy to describe the multilevel processes that generate the data. The unique effects of each group can be specified and estimated through the creation of both within-group (individual) and between-group (school) models. Financial data, however, are still aggregated at the district level and their effects still cannot be effectively estimated in either the individual or school level models.

In this study the continued problem of cross-level aggregation is not overcome completely, but it is certainly reduced. The author demonstrates that the creation of independent variables is often unique to each production function study. This is particularly true in this analysis where there is unique questionnaire-generated data available. In every case, the specification of analytical models and the correlations between variables require careful scrutiny from the reader. A close reading of this work demonstrates the difficulty of compiling results from several studies and drawing

Table 21. Study Characteristics of Holtzapple

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent low income	district	total	school	(-) NS	83	3	Southwest Pennsylvania
				(-) S*	82	5	
				(-) S*	75	8	
		reading		(-) NS	83	3	
				(-) S*	82	5	
				(-) S*	75	8	
		math		(-) NS	83	3	
				(-) S*	82	5	
				(-) S*	75	8	
teacher preparation time	school	total	school	(-) NS	83	3	Southwest Pennsylvania
				(-) NS	82	5	
				(+) S*	75	8	
		reading		(-) NS	83	3	
				(-) NS	82	5	
				(+) NS	75	8	
		math		(-) NS	83	3	
				(-) NS	82	5	
				(+) S*	75	8	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 21 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent African-American pupils	school	total	school	(-) NS	83	3	Southwest Pennsylvania
				(-) NS	82	5	
				(-) NS	75	8	
		reading		(-) S*	83	3	
				(-) S*	82	5	
				(-) NS	75	8	
		math		(-) NS	83	3	
				(-) NS	82	5	
				(-) NS	75	8	
teacher experience	school	total	school	(-) NS	83	3	Southwest Pennsylvania
				(-) NS	82	5	
				(+) NS	75	8	
		reading		(-) NS	83	3	
				(-) S*	82	5	
				(+) NS	75	8	
		math		(+) NS	83	3	
				(-) NS	82	5	
				(+) NS	75	8	

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 21 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
teacher percent African-American	school	total	school	(-) S*	83	3	Southwest Pennsylvania
				(-) NS	82	5	
				(-) NS	75	8	
		reading		(-) NS	83	3	
				(-) NS	82	5	
				(-) NS	75	8	
		math		(-) S*	83	3	
				(-) S*	82	5	
				(-) NS	75	8	
school organization	school	total	school	(-) NS	83	3	Southwest Pennsylvania
				(+) NS	82	5	
				(+) NS	75	8	
		reading		(+) S*	83	3	
				(-) NS	82	5	
				(+) NS	75	8	
		math		(-) NS	86	3	
				(+) NS	82	5	
				(+) NS	75	8	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 21 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
district wealth	school	total	district	(+) S*	83	3	Southwest Pennsylvania
				(+) NS	82	5	
				(+) NS	75	8	
		reading		(+) NS	83	3	
				(+) NS	82	5	
				(+) NS	75	8	
		math		(+) S*	83	3	
				(+) NS	82	5	
				(+) NS	75	8	

Other Comments:
The sample size shows the number of schools without African-American pupils verses the total number of schools. Race was analyzed as a variable in only one data set.
District wealth takes expenditures into account.
The TELLS (Pennsylvania Tests of Essential Learning and Literacy) tests administered in 1989 for grades 3, 5, and 8 in reading, mathemathematics, and total performance scores are the dependent variables. The TELLS test is a diagnostic test used to determine remediation needs. A "cut" score is used for identification that truncates the variance and limits its value in production function analysis.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

conclusions about the effects of school inputs because the specification of independent and dependent variables is substantially unique.

Stiles, Table 22, found significant positive correlations between educational achievement and per pupil expenditures on eleven of eighteen measures of the Missouri Matery Achievement Test (MMAT). There were high correlations between per pupil expenditures, equalized assessed valuation, and average class size (all special teachers were excluded). Positive associations between educational achievement and per pupil expenditures, equalized assessed valuation, the district income factor and the percentage of mobility. Negative correlations between educational achievement and average class size, the percentage of pupils on free or reduced lunch, the percentage of minority pupils, and the percentage of pupils of single parent households were found. The independent variables accounted for little variance in achievement (the highest R^2 = 20.73). SES (as measured in this study) accounted for little variance in achievement among small Missouri schools. The findings indicated that factors affecting achievement in metropolitan schools, particularly SES, have significantly less effect on achievement in rural schools.

The Stiles project selected variable proxies from those available within existing school data collection responsibilities. This practice does not allow for the design of a well-articulated conceptual model and none are offered in the paper. The failure to devise a conceptual model is a methodological flaw in most education production function studies. Without a conceptual model an interpretive analysis is not meaningful (Pedhazur, 1982). Stiles did attempt to formulate a predictive regression model for

Table 22. Study Characteristics of Stiles

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
% minority	school	reading	school	(-) NS (-) S* (+) NS	106	2,3,4 5 6	rural Missouri
		math		(-) NS (+) NS		2,3,4 5,6	
		science		(-) NS (+) NS		3,4,5 6	
		social studies		(-) NS (+) NS		3,4,5 6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 22 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
% single parent home	school	reading	school	(+) NS (-) NS	106	2,5,6 3,4	rural Missouri
		math		(-) NS		2,3,4 5,6	
		science		(-) NS (+) NS		3,4,6 5	
		social studies		(-) NS		3,4,5,6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 22 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
% mobility	school	reading	school	(+) NS (-) NS	106	2,5,6 3,4	rural Missouri
		math		(-) NS (+) NS		2,3,4,6 5	
		science		(+) NS (-) NS		3,4,5 6	
		social studies		(-) NS (+) NS		3,4 5,6	

*p<.05 S = SIGNIFICANT

**p<.01 NS = NONSIGNIFICANT

Table 22 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
% free / reduced lunch	school	reading	school	(+) NS (-) NS	106	2 3,4,5,6	rural Missouri
		math		(+) NS (-) NS		5 2,3,4,6	
		science		(-) NS (+) NS		3,4,5 6	
		social studies		(-) NS		3,4,5,6	
district income	school	reading	school	(-) NS (+) NS	106	2 3,4,5,6	rural Missouri
		math		(-) NS		2,4 3,5,6	
		science		(+) NS (-) NS		3 4,5,6	
		social studies		(+) NS		3,4,5,6	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 22 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
equalized assessed valuation	school	reading	school	(+) NS	106	2,4,5,6	rural Missouri
				(+) S**		3	
		math		(+) NS		2,4,5,6	
				(+) S**		3	
		science		(+) S*		3	
				(+) NS		4,6	
				(-) NS		5	
		social studies		(+) NS		4,5,6	
				(+) S*		3	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 22 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
per-pupil expenditure	school	reading	school	(+) NS	106	2	rural Missouri
				(+) S*		3	
				(+) S**		4,5,6	
		math		(+) NS		2,3,4	
				(+) S*		5	
				(+) S**		6	
		science		(+) NS		3,4,5	
				(+) S**		6	
		social studies		(+) S**		3,6	
				(+) S*		4,5	

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

Table 22 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
class size	school	reading	school	(-) NS (-) S* (+) NS		2,4,5 3 6	rural Missouri
		math		(-) NS (-) S*		2,4,5,6 3	
		science		(-) NS (+) NS		2,3,4,6 5	
		social studies		(-) NS (-) S*		2,4,5,6 3	

Other Comments:
Missouri Mastery Achievement Tests (MMAT) are criterion-referenced tests. The chart reports correlation data. A sample of 106 elementary schools containing grades 2-6 from 94 districts with enrollments between 300 and 1,800 was selected from a population of 222 elementary schools within 193 school divisions containing high schools.

*p<.05 S = SIGNIFICANT
**p<.01 NS = NONSIGNIFICANT

evaluating the performance of pupils within the individual school setting for sample participants. This is a more satisfactory use of the regression coefficients since they are generally sample specific. Unfortunately, parts of the data are aggregated at the district level and parts are aggregated at the school level which inhibits a wholly satisfactory depiction of school variable relationships.

The match between the taught and measured curricula is more completely assured with the utilization of a criterion-referenced test such as the MMAT. Nationally normed standardized tests are the most commonly used output instruments. Scope and sequence variations between these tests and state and local curricula generally confound the accuracy of production function analyses.

Ferguson, Table 23, evaluated data in 890 of more than 1000 districts in Texas that included 2.4 million students, nearly five times the number of students represented in the Coleman report, using two stage least squares regression procedures. Excluded districts were generally small and lacking important data, although Dallas and Houston were omitted because their size would have overwhelmed results from other districts. Primary data came from the Texas Education Agency, as well as other state agencies, school districts, local governments, and a special tabulation of the 1980 U.S. Census that provided data by school district. Particularly valuable information on teachers was available from the certification literacy test initiated in 1986 for all teachers and administrators, the Texas Examination of Current Administrators and Teachers (TECAT). The analysis used passing rates that were compiled separately for primary and secondary school teachers. Average scores were not available. Student achievement

Table 23. Study Characteristics of Ferguson

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
TECAT score	district	reading	district	(+) S**	838 districts	composite of grades 3, 5, 7 9, 11	Texas
teacher experience	district	reading	district	(+) S**	838 districts	composite of grades 3, 5, 7, 9, 11	Texas
per cent of teachers with Master's degrees	district	reading	district	(+) S**	838 districts	composite of grades 3, 5, 7, 9, 11	Texas
teacher salaries	district	reading	district	(+) S*	838 districts	composite of grades 3, 5, 7, 9, 11	Texas

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 23 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
percent of African-American pupils	district	reading	district	(-) S*	838 districts	composite of grades 3, 5, 7, 9, 11	Texas
percent of Hispanic pupils	district	reading	district	(-) S*	838 districts	composite of grades 3, 5, 7, 9, 11	Texas
teachers' language skill	district	reading	district	(+) S**	838 districts	composite of grades 3, 5, 7, 9, 11	Texas
percent of parents with some college education	district	reading	district	(+) S**	838 districts	composite of grades 3, 5, 7, 9, 11	Texas

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

Table 23 (continued)

Input Variables		Output Variables		Direction of Effect	Sample Size	Grade Level	Location
Measure	Level of Aggregation	Measure	Level of Aggregation				
female head of household	district	reading	district	(-) S*	838 districts	composite of grades 3, 5, 7, 9, 11	Texas
number of pupils per teacher	district	reading	district	(-) NS	838 districts	composite of grades 3, 5, 7, 9, 11	Texas

Comments:
The Texas Examination of Current Administrators and Teachers (TECAT) was a literacy test first administered to all teachers and administrators in 1986.
The dependent reading variable identified in this chart came from the Texas Educational Assessment of Minimum Skills (TEAMS) exams administered to the same group of pupils in the 1985-86, 1987-88, and 1989-90 school years.
The sample size used for composite grade scores, 838, represented the smallest number of districts used in any of the equations in the study. Individual grade levels ranged from 857 for the eleventh grade to 890 for the first grade.
Teacher supply equations used data from 887 districts.

*$p < .05$ S = SIGNIFICANT
**$p < .01$ NS = NONSIGNIFICANT

was measured by mean district scores from the Texas Educational Assessment of Minimum Skills (TEAMS) exams that were administered to grades 3, 5, 7, 9, and 11. Data for grade 1 were not included in the composite grade totals.

This study produced a rich body of findings from one of the most complete data sets ever compiled. The author found evidence that instructional expenditures have significant effects on student achievement when they resulted in the hiring of teachers with strong literacy skills, hired more teachers (where the ratio of teachers to students exceeded 18), retained experienced teachers and attracted more teachers with advanced training when socioeconomic (SES) measures were controlled. Consistent with most previous studies, some SES factors also had significant effects on achievement. The percentage of parents with some college education had a very significant positive effect in predicting student performance, even more important than family income. Female headship of families (a proxy for single parent families) demonstrated a statistically significant and negative effect for grades 1-9 and tended to be more important than the percentage of children living in poverty. The percent of Hispanic students was consistently significant and negative, but the percent of African-American students had highly significant effects for only grades nine and eleven. Location (city, suburb, town, rural town, non-metropolitan city) proved significant only for districts with extreme poverty rates that were geographically located along the United States-Mexican border. Ferguson's findings suggested that differences in the quality of schooling accounted for 25% to 33% of the variation in student test scores due almost entirely to teacher quality measured by TECAT. Intrigued by the high correlation coefficient between this measure of teacher quality and high SES districts, he also examined issues related to the supply of teachers and suggested mechanisms to improve the quality of personnel in low SES districts.

Despite the sophisticated statistical methodology, the richness of some data, and the intellectual quality of the analysis, this study lacked important information to meet all the criteria specified by the conceptual model. Indeed, Ferguson noted "the surprising scarcity of data appropriate for establishing the relative importance of various schooling inputs" (p. 465). No measurement for innate abilities was available, although the author used a value-added analysis as a substitute approach. School district averages for SES were used rather than measures for families with children in school. Family and community background measures were not separated for entry into the equations, because the data were not collected in a way that would permit disaggregation. Students per teacher measures were used because data persuant to average class size were unavailable. This measure distorted the actual classroom count by including guidance counselors, special education teachers, or music, art, and physical education teachers who may act as itinerants in primary/elementary schools. Ferguson defended his use of 1980 Census data for SES measures with 1985-86 test data. Undoubtedly, some districts experienced demographic changes during the period, although it is plausible that the large number of districts limited the effect of these changes on the results. A more significant problem lay in the application of this data to student test performance in later years. By using test data to provide a value-added analysis incorporating test scores for the same group from

1985-86, 1987-88, and 1989-90 the gap between demographic and test data widened to a decade.

All the data for this study were aggregated at the district level. Although Ferguson expressed a preference for data on individual pupils and teachers for some analyses, he felt that any examination of disaggregated data would add interesting details, but not change the essential findings of the study. He argued that the "findings...disclose a systematic and internally consistent story whose coherence and plausibility indicate strongly that aggregation bias is not a serious problem in this study" (p. 471). These caveats do not change the reality that this analysis, like other education production function studies, face serious data deficits. The interpretation of the results of this study may not be distorted by biases of aggregation or variable and model specification, but the sensitivity of the author to these issues should caution us that even the best study must overcome serious methodological problems if it is to offer solid answers about the factors the influence student achievement.

CHAPTER V

FINDINGS

An examination of the studies presented in this paper reveals that education production function studies do show that financial resources can influence the academic performance of pupils in K-12 education when certain assumptions are met. Patterns in the relationships between school expenditures and achievement, however, are difficult to identify.

No study examined in this paper fully met the criteria of the conceptual model. The primary model deficit in these analyses is the lack of a measure for initial endowments or innate abilities. The extensive use of cross-sectional data limited the opportunity to examine gains over time. Some studies, however, used test scores from the first grade or from a year preceding the scores used in the analysis to control for past instruction and ability in a variation of the conceptual model that applies a value-added approach. Yet, gain scores face validity challenges in any analysis and they must be related to a human growth curve to provide meaningful explanatory value. No study attempted to meet these demands.

Dependent variables evaluated in these studies differed in three primary areas:

1. skills--composite totals, composite of math and reading or language arts, reading only, math only, science only, or measures from several areas listed separately;
2. tests--ITBS, CTBS-U, Stanford Achievement Tests, various state criterion-related tests;

3. measures--mean raw scores, mean local curve equivalent (LCE) scores, mean standard scores; pass-fail scores; and grade equivalent (GE) scores.

Variations in these three areas that result from different state and local decisions make replication of any complete study difficult outside of a limited political-geographic unit. Indeed, data from each state tend to be idiosyncratic because of differences in definitions of data, funding structures, the relationship between political units and school districts, and unique size and demographic distributions within districts in each state (see Salmon, et.al, 1988 on varieties of fiance programs).

Political subdivisions in some states are identical to school districts. This can allow for the use of socioeconomic data from census collections. In states where cities and counties are dissected by several school districts of various sizes and constituencies, county or city level data may be grossly inaccurate for a particular school district. Distribution patterns for ethnic and income level populations are not likely to be random. If census data representative of the county or city as a whole are used for the analysis or adjusted to reflect the relative size of the district's population, these data can confound the results of an education production function study.

The skills evaluated as measures of achievement can have an effect on the results of the analysis. Reading skills tend to respond more slowly to instruction and reflect the influence of socioeconomic factors to a greater degree than math skills. Math scores tend to respond to instruction more readily than reading scores. Yet, both these skills and others were found to be positively influenced by financial resources in some studies.

State-developed criterion-referenced tests more closely relate to what students are taught than norm-referenced tests since they reflect a state-mandated curriculum. It is not surprising that these tests are more prominent in studies that report positive findings. Norm-referenced tests, however, were also dependent variables in studies that produced positive results. No clear pattern of effect for criterion-referenced tests over standardized tests emerged from the analysis.

Test scores for every individual contain errors; any test score deviates from a true score. Aggregation multiplies the errors. The higher the level of aggregation the greater the error in dependent variables. Since most studies tend to rely upon one set of tests and aggregate to the district level, the errors in individual scores are magnified. Moreover, relating the individual student performance to district level investments in education can be misleading in cases where rapid population shifts occur. This can be a problem with panel data as pupil transfers into and out of the district alter the demographic composition or educational background of test groups.

The level of statistical significance for most analyses is set at .05 and below. Some studies used higher or lower levels, Brock for example chose .10. Such a decision is within the power of the researcher, but it clearly affects the summary conclusions that are frequently the reference point for discussions about the issue. Another example, Hanushek, only reported significance and direction of significance in his charts while the level set by the researcher remained unlisted.

Aggregation affects sample size. Studies report the number of students as if this has some importance in the analysis when aggregating to the district level.

Unfortunately, district level studies are frequently reduced to small sample sizes. Meaningful conclusions cannot be made from such small samples. Ferguson is an exception to this condition. His study is unique in the number of districts and the quality of his data. One study, however, cannot carry the burden for thirty years of analyses. Replication must be possible, and it must be done on a regular basis. Without it education production functions are entirely hypothetical.

Ultimately, the inability to randomize the sample of pupils, schools, or districts violates the assumptions of basic statistical methodology. Nonrandom samples are not generalizable to the larger population. Comparisons between studies, such as those attempted by Hanushek, are simply meaningless.

Systematic evidence from education production function studies does not exist to prove or disprove that financial resources influence educational success, because there is no standardized, systematic analysis. Studies that find positive and significant results violate the conceptual model just as those that do not find such results.

CHAPTER VI

STANDARDS OF ADEQUACY FOR THE EDUCATION MODEL

Quality Variations of Inputs

An ideal model requires that variables be manipulable and predictable. The previous reference to model validity remains an overriding concern for the usefulness of the production function analysis in education. A means must be found to confirm the validity of the model's use in particular settings. There may be a different production function for every individual, school, or district (Monk, 1990). The compelling need is to identify and use it. Quality variations in inputs refer in part to differences of type. Lau (1979) identified variables like Cobb and Douglas--fixed and variable. Some inputs are fixed for educators over long periods of time, such as basic facilities and type of school (elementary, middle, or high and public or private); some are variable in the long run but fixed in the short run, such as teacher characteristics; some are variable in the short run, such as teacher and student time. No one manager controls all these inputs, yet every actor influences their effect on the production function. A means must be found to validate the model in a given setting. To do so requires greater attention to overcoming difficulties in the identification and quantification of inputs.

Inability to Measure Skill Differences in Teachers

The inability to measure skill differences between teachers who share similar quantifiable attributes such as level of education, years of experience, or certification credentials, is a deficiency in education research and, thus, production function analysis. Differences in teaching styles, strategies, attitudes, and behavior of teachers are likely to be important factors in student achievement. Significantly different amounts of learning do occur in different classrooms in the same school and between different schools. Clearly, teachers and schools can have significant effects on student achievement. Production function studies are simply unable to capture the behaviors that produce achievement (Hanushek, 1972; Murnane, 1975; Armor, et al., 1976; Murnane & Phillips, 1981).

Murnane and Nelson (1984) argue against the use of production function analysis in education because of the notion of teacher differences. Production function analysis is based on a standardized process derived from fully articulated and detailed production inputs and techniques that allow very little variation. Teaching is characterized by techniques that are essentially situation specific and idiosyncratic. Teachers are constantly making production decisions in the classroom about methods, procedures, and materials. This promotes considerable variation in the teaching process. Teaching inherently involves constant experimentation to identify those techniques that will increase student performance in the classroom. Education inputs and techniques gradually evolve as teachers reshape and modify instruction. Research is normally conducted away from the setting of schools and does not capture the realities of the process of constant adjustment and modification necessary to meet the needs of individuals and classroom groups.

Suggestion for Validation

The combination of the strategies described by two studies offers the best chance to validate production function studies. Since input variables lie outside the ability of one person to control, they must be analyzed in conditions that allow for the observation of limited changes to inputs under careful scrutiny. Thompson and Correa's three year analysis of a group of very stable relationships offers a good beginning. Students in this cohort could be followed (in middle and high school) to view the effects of variable changes on their performance. This kind of painstaking analysis pursued over time is the best way to attack the validity dilemma. The reduction of sample size introduced by these changes becomes another problem that must be dealt with at the beginning of the project by the selection of a large sample. The type of longitudinal study represented by Card and Krueger's (1992) evaluation of gains to education is the second step to be applied to production function analysis. If successful predictions from a model developed with the Thompson and Correa-type study can be made of the performances of students from a Card and Krueger-type study, validation is possible.

Objectives Specification and Multiple Outputs

Objectives specification can best be described as the identification of what is taught in a classroom and measured on a test. A perfect correlation between the two is highly desirable, but not always achieved. Criterion-referenced tests attempt to measure these relationships directly. Standardized tests measure general abilities that are part of the instructional process, but are normed nationally for the entire student population.

They measure an important part, but certainly not all, of what is taught. Standardized tests do not drive the curriculum, and their nearly exclusive use as output measures in production function analysis fails to measure important objectives of schooling.

Spencer and Wiley (1981) described the relationship between goals and resources. Schools may differ significantly in the goals they set and direct resources accordingly. Educational goals are shaped by political and religious beliefs, as well as, economic and social class. Students may spend more time on drama and music in one school and more on computer drill and practice at another. The educational experience can be very different in these settings, irrespective of standardized testing or financial support. In high school, students may follow a vocational or academic track because of peer or family/community relationships unaffected by academic abilities. Standardized test scores cannot measure the educational values reflected in these decisions, but these values are reflected in test performance.

The choices that parents, communities, and educators make are in the goals of the educational program and find expression in the curriculum as behavioral objectives. Multicultural education is an example. Gollnick (1980) described five goals for an instructional program: 1) Promoting the strength and value of cultural diversity; 2) Promoting human rights and respect for those who are different from oneself; 3) Promoting alternative life choices for people; 4) Promoting social justice and equal opportunity for all people; 5) Promoting equity in the distribution of power among groups. The Iowa Test of Basic Skills, for example, will not provide achievement measures for these objectives.

The measurement of diverse objectives is the major issue of specification on the output side. Since 1989 the Commonwealth of Virginia, for example, has begun to collect and publish broad-based output measures in the Outcome Accountability Project (OAP). The measures relate to seven state-wide objectives: preparing pupils for college, preparing pupils for work, increasing the graduation rate, increasing special education pupils' living skills and opportunities, educating elementary school pupils, educating middle school pupils, and educating secondary school pupils. The proxies used for these goals include: measures of access to programs, levels of participation, and standardized test performance. The development of such multiple output measures related to a variety of program objectives may result in the development of a useful production function model. However, the measures must be redesigned to provide interval level data to accommodate statistical analysis.

Level of Analysis

The findings of empirical studies have been challenged for mixing levels of analysis between district, school, classroom, and student. Model adequacy requires uniformity of level. The important question to address, however, is the appropriate level for an accurate, policy-useful model. Some researchers have concluded "there probably aren't many policy manipulable predictors of academic achievement much above the level of the individual classroom" (Hereford & Keith, 1991, p. 27). Since this is the locus of direct school inputs to each student's education, it is not a surprising conclusion. It would, however, render most research in the area invalid for policy purposes.

The collection of data at the classroom level is the direction research activities must take to develop a useable production function model. That is not to say that other levels of analysis cannot be valid. It only appears that the sequence of development is best pursued inductively. The time and resources such a process requires are rarely available to school-level practitioners. Only a district, state, or federal effort is likely to produce the necessary data over the period of time necessary for model validation.

Apportioning Value Added Performance

Determining the contribution of school factors to student achievement became an important research objective in the wake of the Coleman Report. The recognition of the need for a measure of ability or achievement as a pretest score is also a vital concern in the development of an adequate model of education production. Lau (1979) and others have acknowledged the "uninformative" nature of cross-sectional studies that lack such measures. Bowles (1970) described achievement scores as measures of gross output. The goal was "to estimate the relationship between school inputs and *net* output, or value added" (p. 26).

Burkhead (1967) used IQ scores for the 10th grade and achievement scores for 12th grade reading to estimate the gains of student performance in Chicago and different levels of similar data for Atlanta. He acknowledged his preference for achievement test scores for both measures from the same test, but, the data were unavailable. Hanushek (1972) incorporated measures of initial endowment into his model, but performed analyses which lacked these data. Bowles (1970) cited evidence from Bloom that achievement and intelligence scores were correlated at .85. This close relationship could be interpreted in several ways that could under or over estimate the effect of schooling. He expressed a preference for administering an achievement test in the first grade to be used for this purpose. The Commonwealth of Virginia, for example, used to administer the Cognitive Abilities test in October of the first grade year. Measures that allow the assessment of value added performance are generally regarded as important components of an adequate model. The lack of such data seriously impairs the usefulness of empirical studies.

Standardized Tests and Data Collection

At this time there are no national standardized tests, only a variety of tests that have national norms. The debate over national testing now being waged suggests that the issue remains volatile. Indeed, no impetus at any level of government exists to collect data that would meet the needs of production function studies. Even Ferguson (1991), with his data rich material, could construct a study at the district level only. The development of an adequate model requires the purposeful collection of data that has direct relevance to a conceptual production function design. A national testing program could provide the impetus for such a collection process. Yet, a strong movement among educators challenges the adequacy of quantitative measures of student achievement. The development of a broad-based qualitative methodology for student evaluation through portfolio building has begun to spread across the country. This approach is based, partially, on dissatisfaction with purely statistical appraisals. For this reason alone, any effort to collect national data for production function analysis would probably face vigorous political opposition.

The Education Model in Policy Making

The Coleman Report

Without elaborating on the details of the Coleman Report, consider the example it perpetuates for policy makers. The formulation of policy on the basis of a half-formed conceptual methodology can result in faulty decisions and undermine support for social science research. Indeed, Coleman himself eventually recanted some of the conclusions contained in the report. Using the findings from education production studies to determine important policy issues can multiply the errors of the EEO Report. Dissatisfaction with policies issuing from the Coleman era has contributed to the decline of data collection and research projects on a magnitude that would provide opportunities to validate the education production model. Premature advocacy of production function conclusions can lead to reactions such as that expressed by Christopher Edley, Jr. (1991):

> However imperialistic lawyers can be in offering their services and habits of mind in the solution of all problems, economists are even more dangerous with implicit claims that their grossly simplified models should displace the instincts and experiences of professionals, such as educators, who have worked for decades to understand the ingredients of progress. Crucial research must focus not on the empirical analysis of aggregate input-output models, but on the more conventional, less tidy, applied problem of program evaluation and replication. That is how social science can best serve struggling educators and advocates, who ought not to be diverted to rebutting and perfecting flawed economic models (p. 296).

Money and the Courts

The effort to provide equal educational opportunities for children in this country was not achieved as a consequence of the Coleman Report and subsequent legislation. The fundamental proposition that the quality of a child's education should *not* be determined by the wealth of the community in which the child lives translated into a quest for equal per pupil funding (see Alexander, 1991, 1982 for discussions on funding equity). Because legislatures did not resolve the issue, it entered the federal and state courts. In 1973 with San Antonio v. Rodriguez, the Supreme Court referred the educational funding challenge to the states arguing that education is a state responsibility under the U.S. Constitution. State courts have been the locus of litigation since that time.

The erratic findings of production function studies have been used as a political weapon for those who would perpetuate the inequities that exist in school funding. Hanushek (1991) and others have claimed that "money does not matter" based on some selected analyses. Kozol (1991) quotes the Wall Street Journal that "money doesn't buy better education." He notes that "the Journal does not tell its readers that the current average figure masks disparities between the schools that spend above $12,000 . . . and the ones that spend less than $3,000. Many of the poorest schools today spend less than the average district spent ten years ago" (p. 133). Those who read Kozol's book are likely to conclude that the production model misses some very important details. Many production function studies, however, do cite ways in which money makes a difference in the education of children. Ferguson (1991) is the most recent example. The quantity and quality of his data lend far more credibility to his findings than that of most earlier studies.

The desire of the courts for information with which to make informed decisions is understandable. The opportunity for researchers to influence policy is compelling. However, decisions based on the production function model in its current state can be accurate only by chance. It's explanatory power is problematic, and its predictive ability is without validation.

"Efficiency" and the State of the Art

> For while capital may be "productive," it does not follow that the capitalist always is. Capital would still be "productive" even though its ownership were changed. Nor does it follow that the uses to which the capitalists put the income which they receive are on the whole socially the best. One may therefore be still a supporter of socialism, communism, or individualism and still square his social philosophy with the theory of production which we have developed (Cobb & Douglas, 1928, p. 165).

Levin (1974, 1976) argues that schools are inherently inefficient and that production function results from cross-section results of pupil performance do not represent what could be accomplished, only what is accomplished. Inefficiency comes from three sources:

1. Schools are not operating on the "production frontier"--at the point where the best technology is employed.
2. Because of the technology used by schools, resources are not combined in a manner that will produce the maximum output from the available inputs.

3. Schools are not likely to respond to the true desires of society relative to the proper mix of educational outputs.

These conclusions suggest that data from current school analyses are inappropriate for generating information that will improve resource allocation in schools.

Hanushek (1976, 1979, 1986) states that the issue of inefficiency is clouded by the confusion of different concepts of efficiency; economic efficiency ("correct choice of input mix given the prices of inputs and the production function") and technical efficiency ("operating on the production frontier," that is, using the best available techniques in the production of education). Two arguments are used to support the idea of inefficiency in the public schools:

1. Educational decision makers lack incentives to increase student achievement or reduce costs.
2. They do not have an adequate understanding of the educational production process.

Hanushek denies that these arguments negate the value of production function studies. The first claim suggests inefficiency in the system, but it does not prove that schools are "off the production frontier unless resources are also wantonly squandered" (1979, p. 370). The second claim refers to organizational processes that can be evaluated by the production function methodology. Empirical production function studies are based on existing practice. They can provide information about improvements in the existing framework of conditions. The introduction of improved technologies lies outside empirical analysis.

Hanushek (1972) identified two criteria for assessing school efficiency: knowledge of the relationship between inputs and outputs of the educational process and a decision rule to connect costs of various inputs to the process with their educational outputs. In this paper many of the problems encountered in production function analysis have been examined. Researchers, including Hanushek, agree that the actual relationship between inputs and outputs in education is unknown. The decision rule cannot connect costs to the model. To do so is to compound our ignorance since the model cannot be complete. Hanushek (1992, 1989, 1986, 1981) has also argued that the inconsistent results involving so many studies demonstrate that schools are inefficient. What this confusion really demonstrates most conclusively is the insufficiency of necessary data and a model incapable of validation. Schools may be inefficient. Unfortunately, existing education production function studies cannot really tell us this. Cobb and Douglas maintained that the industry model is consistent across all ideological forms of economic organization. The education model is inconsistent in every setting. The weight of the evidence from this study establishes the unreliability of the model and its findings. The process of schooling is such a complex process that we need a much more complex model. What is described as inefficiency may be corrective action within the system that makes it functional.

The production function study is an attractive research tool, but the promise of what it can provide for policy makers remains unfulfilled. If refinements in the collection of data and empirical applications continue, the development of a validated model may be possible.

References

Alexander, K. (1991, November). Financing the nation's schools: A perspective on fiscal effort and educational need. Paper presented as testimony to the Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor, U.S. House of Representatives, Washington, DC.

Alexander, K. (1982). Concepts of equity. In W. W. McMahon & T. G. Geske (Eds.), Financing education: Overcoming inefficiency and inequity (pp. 193-214). Chicago: University of Illinois Press.

Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., & Pincus, J. (1971). How effective is schooling?: A critical review and synthesis of research findings. Santa Monica, CA: The Rand Corporation.

Baker, K. (1991, April). Yes, throw money at schools. Phi Delta Kappan, 72(8), 628-631.

Barro, S. M. (1989, Spring). Fund distribution issues in school finance: Priorities for the next round of research. Educational Evaluation and Policy Analysis, 11(1), 17-30.

Berliner, D. C. (1993). Educational reform in an era of disinformation. Education policy archives: An electronic journal. EDPOLYAR@ASUACAD.BITNET.

Bloom, B. S. (1964). Stability and change in human characteristics. New York: John Wiley & Sons.

Boardman, A. E., Davis, O. A., & Sanday, P. R. A simultaneous equations model of the educational process: The Coleman data revisited with an emphasis on achievement. Washington, D.C.: American Statistical Association. (ERIC Reproduction Service No. ED 097 404)

Booms, B. H. (1969). Empirical estimates of secondary education production functions. (Doctoral dissertation, University of Pittsburgh, 1968). Dissertation Abstracts International, 69-4090.

Borgatta, E. F., & Jackson, D. J. (1980). Aggregation data analysis: An overview. In E. F. Borgatta & D. J. Jackson (Eds.), Aggregate data: Analysis and interpretation. Beverly Hills, CA: Sage Publications.

Bosworth, D. L. (1976). Production functions: A theoretical and empirical study. Westmead, England: Saxon House, D. C. Heath.

Bowles, S. S. (1970). Towards an educational production function. In W. L. Hansen, (Ed.), Education, income, and human Capital. New York: National Bureau of Economic Research.

Bowles, S. S., & Levin, H. M. (1968). The determinants of scholastic achievement--an appraisal of some recent evidence. Journal of Human Resources, 3, 393-400.

Bridge, G. R., Judd, C. M., & Moock, P. R. (1979). The determinants of educational outcomes: The impact of families, peers, teachers, and schools. Cambridge, MA: Ballinger Publishing Co.

Bronfenbrenner, M., & Douglas, P. H. (1939, December). Cross-section studies in the Cobb-Douglas function. The Journal of Political Economy, 47(6), 761-785.

Brock, J. H. (1986). A study of the relationship of pupil achievement test scores in reading and mathematics to pupil expenditures and selected district socioeconomic variables. (Doctoral dissertation, University of Kentucky, 1986). Dissertation Abstracts Internation, 87-15918.

Burkhead, J. (1967). Input and output in large-city high schools. Syracuse, NY: Syracuse University Press.

California state testing program, 169-70: A district by district analysis of test scores and other school factors. Sacramento, CA: Office of Program Evaluation, California State Department of Education. (ERIC Document Reproduction Services No. ED077 991)

Cain, G. G., & Watts, H. W. (1970). Problems in making policy inferences from the Coleman Report. American Socioloogical Review, 35, 228-242.

Campbell, D. T. (1963). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris, Problems in measuring change (pp. 212-245). Madison, WI: The University of Wisconsin Press.

Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. Journal of Political Economy, 100(1), 1-40.

Childs, T. S., & Shakeshaft, C. (1986, Fall). A meta-analysis of research on the relationship between educational expenditures and student achievement. Journal of Education Finance, 12, 249-263.

Chubb, J. E., & Hanushek, E. A. (1990). Reforming education reform. In H. Aron (Ed.), Policy for the nineties (pp. 213-247). Washington: Brookings Institute.

Clausen, A. F. (1989). A study of the relationship between scholastic achievement and local financial effort in support of schools (Doctoral dissertation, Auburn University, 1989). Dissertation Abstracts International.

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

Coffey, J. H. (1991). Student achievement, community socioeconomic charactersitics, and categories of educational expenditure in Connecticut Public Schools. (Doctoral dissertation, University of Bridgeport, 1991). Dissertation Abstracts International, 92-06065.

Cohn, E., & Geske, T. G. (1990). The economics of education (3rd ed.). Oxford: Pergamon Press.

Coleman, J. S., et al. (1966). Equality of educational opportunity. Washington, DC: Government Printing Office.

Commonwealth of Virginia. (1986, October). Excellence in education: A plan for Virginia's future. The Report of the Governor's Commission on Excellence in Education. Richmond: Author.

Connors, E. T. (1982). Educational funding and pupil achievement: The relationship between the two variables. Final report. Richmond: Virginia State Department of Education. (ERIC Document Reproduction Service No. ED 242 085)

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons.

Deposition of Eric A. Hanushek. (July 21, 1992). In the matter of Alabama Coalition for Equity, et al. v. Guy Hunt, et al. Circuit Court for Montgomery, Alabama. CV-90-883-R. Montgomery, Alabama: Parham & Associates.

Dolan, R. C., & Schmidt, R. M. (1987). Assessing the impact of expenditure on achievement: Some methodological and policy considerations. Economics of Education Review, 6(3), 285-299.

Dunkelbrerger, J. E. & Soderberg, M. A. (1980). Nonmetro-metro differences in student achievement scores. Washington, D.C.: Cooperative State Research Service, Department of Agriculture. (ERIC Document Reproduction Service No. ED 190 339)

Dunnell, J. P. (1970). Input and output analysis of suburban elementary school districts. (Doctoral dissertation, Illinois State University, 1970). Dissertation Abstracts International, 70-18107.

Edley, C. F., Jr. (1991, Summer). Lawyers and education reform. Harvard Journal on Legislation, 28(2), 293-305.

Educational Testing Service, Policy Information Center. (1991). The state of inequality. Princeton, NJ: Author.

Erickson, E. H. (1963). Childhood and Society, 2nd Edition. New York: W.W. Norton.

Florio, D. H. (1980, April 8). Education research and the Congress: Unmet potential or unused resource? Paper presented to the American Education Research Association meeting, Boston, MA.

Ferguson, R. F. (1991, Summer). Paying for public education: New evidence on how and why money matters. Harvard Journal on Legislation, 28(2), 465-498.

Fortune, J. C. (1992). Rebuttal to the Deposition of Eric A. Hanushek. Unpublished document. Blacksburg, VA: Virginia Tech.

Gilman, D. A. (1988, Summer). A critical analysis of the use of production funtions in school finance. Planning & Changing. 19(2), 79-85.

Glasman, N. S., & Biniaminov, I. (1981, Winter). Input-output analyses of schools. Review of Educational Research, 51(4), 509-539.

Gollnick, D. M., Baptiste, H. P., & Baptiste, M. L. (1980). Multicultural teacher education. Washington, DC: Commission on Multicultural Education, American Association of Colleges for Teacher Education.

Gray, P. H. (1958). Theory and evidence of imprinting in human infants. Journal of Psychology, 46, 155-166.

Harris, C. W. (Ed.). (1963). Problems in measuring change. Madison, WI: The University of Wisconsin Press.

Hanes, C. E. (1976). Toward greater educational productivity. In K. Alexander & K. F. Jordan, Educational need in the public economy. Gainesville: The University of Florida Presses.

Hanushek, E. A., & Kain, J. F. (1972). On the value of equality of educational opportunity as a guide to public policy. In F. Mosteller & D. P. Moynihan (Eds.), On equality of educational opportunity. New York: Vintage Books.

Hanushek, E. A. (1972). Education and race: An analysis of the education production process. Lexington, MA: D. C. Heath.

Hanushek, E. A., & Jackson, J. E. (1977). Statistical methods for social scientists. New York: Academic Press.

Hanushek, E. A. (1978). Conceptual and empirical issues in the estimation of educational production functions. The Journal of Human Resources, XIV(3), 351-388.

Hanushek, E. A. (1981). Throwing money at schools. Journal of Policy Analysis and Management, 1(1), 19-41.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. Journal of Economic Literature, 24, 1141-1177.

Hanushek, E. A. (1987). Education production functions. In G. Psacharopoulos (Ed.), Economics of education: Research and studies. Elmsford, NY: Pergamon Press.

Hanushek, E. A. (1989, May). The impact of differential expenditures on school performance. Educational Researcher, 18(4), 45-51.

Hanushek, E. A. (1991, Summer). When school finance "reform" may not be good policy. Harvard Journal on Legislation, 28(2), 423-456.

Harbison, R. W., & Hanushek, E. A. (1992). Educational performance of the poor: Lessons from rural northeast Brazil. New York: Oxford University Press/The World Bank.

Hartman, W. T. (1988, Spring). District spending disparities: What do the dollars buy? Journal of Education Finance, 13, 436-459.

Henderson, J. M., & Quandt, R. E. (1971). Microeconomic theory: A mathematical approach. New York: McGraw-Hill.

Hereford, K. T., & Keith, T. Z. (1991). Effects of local financial effort on school district achievement. Blacksburg, VA: Virginia Tech.

Hokanson, S. L. (1984). A causal model: Effects of school expenditure patterns on student achievement. (Doctoral dissertation, Florida State University, 1984). Dissertation Abstracts International, 84-16706.

Holtzapple, E. A. (1991). Schools, student characteristics and student achievement: A Hierearchical Linear Modelling test in Southwestern Pennsylvania. (Doctoral dissertation, University of Pittsburgh, 1991). Dissertation Abstracts International, 92-18873.

Howell, D. C. (1987). Statistical methods for psychology (2nd ed.). Boston: PWS-Kent Publishing.

Huck, S. W., Cormier, W. H., & Bounds, W. G. (1974). Reading statistics and research. New York: Harper Collins, Publishers.

Hughes, M. F. (1991, Winter). A comparison of social, economic, and educational variables in high and low expenditure school districts of Virginia, 1988-89. Journal of Education Finance, 16, 388-397.

Jewell, R. M. (1971). The relationship of selected fiscally related variables and pupil achievement. (Doctoral dissertation, Duke University, 1971). Dissertation Abstracts International, 72-16987.

Johns, R. L., Morphet, E. L., & Alexander, K. (1983). The economics & financing of education (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kozol, J. (1991). Savage inequalities: Children in America's schools. New York: Crown.

Lau, L. J. (1979). Educational production functions. In Economic dimensions of education. Washington, DC: The National Academy of Education.

LaValley, T. W. (1990). Input-output analysis of educational expenditures on student achievement. (Doctoral dissertation, University of Oklahoma, 1990). Dissertation Abstracts International, 90-29780.

Lee, C. M.-S., Aron, I-M. W., & Aron, R. H. (1988, Spring). An analysis of Michigan educational assessment program (MEAP) scores and school districts revenues and expenditures. Journal of Education Finance, 13, 496-511.

Lopus, J. S. (1990). Expenditure effects on learning and achievement in high school economics classes. (Doctoral dissertation, University of California, Davis, 1990). Dissertation Abstracts International, 91-10648.

McDowell, G., Elias, C., & Driscoll, P. (1992). Paying for schooling in Virginia: A citizen's guide to school finance. Blacksburg, VA: Virginia Cooperative Extension.

MacPhail-Wilcox, B., & King, R. A. (1986, Spring). Resource allocation studies: Implications for school improvement and school finance research. Journal of Education Finance, 11, 416-432.

MacPhail-Wilcox, B., & King, R. A. (1986, Fall). Production-functions in the context of educational reform. Journal of Educational Finance, 12, 191-222.

Michelson, S. H. (1970). The existentialist reality of educational production functions (their existence has preceded their essence). Washington, D.C.: National Center for Educational Research and Development. (ERIC Document Reproduction Service No. ED 073 009)

Monk, D. H. (1981, Summer). Toward a multilevel perspective on the allocation of educational resources. Review of Educational Research, 51(2), 215-236.

Monk, D. H. (1989, Spring). The education production function: Its evolving role in policy analysis. Educational Evaluation and Policy Analysis, 11(1), 31-45.

Monk, D. H. (1990). Educational finance: An economic approach. New York: McGraw-Hill Publishing.

Monk, D. H. (Winter, 1992). Education productivity research: An update and assessment of its role in education finance reform. Educational Evaluation and Policy Analysis, 14(4), 307-332.

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

Murnane, R. J. (1975). The impact of school resources on the learning of inner city children. Cambridge, MA: Ballinger.

Murnane, R. J. (1991, Summer). Interpreting the evidence on "does money matter?" Harvard Journal on Legislation, 28(2), 457-464.

Murnane, R. J., & Nelson, R. R. (1984). Production and innovation when techniques are tacit. Journal of Economic Behavior and Organization, 5, 353-373.

Murnane, R. J. & Phillips, B. R. (1978). effective teachers of inner city children: Who they are and what they do. Washinton, D.C.: Finance and Productivity Group,k National Institute of Education. (ERIC Document Reproduction Service No. ED 166 348)

Nichols, R. C. (April, 1976). School effects on achievement. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California. (ERIC Document Reproduction Service No. ED 126 104)

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: CBS College Publishing.

Pelton, M. D. (1966). A national analysis of educational quality as measured by the Educational Characteristics Criterion (ECC), achievement, and selected cost factors. (Doctoral dissertation, Michigan State University, 1966). Dissertation Abstracts International, 67-1671.

Phillips, J. M., & Marble, R. P. (1986). Farmer education and efficiency: A frontier production function approach. Economics of Education Review, 5(3), 257-64.

Pollanen, S. (1991). Equity of educational achievement and school effectiveness. Rochester, NY: Greece Central School District, Research and Program Evaluation. (ERIC Document Reproduction Service No. ED 333 021)

Price, A. H., Strickland, D. C., & Fortune, J. C. (1993). A crtitical review of Walberg and Hanushek's contribution to school equity issues. Proceedings. Atlanta, GA: The 11th Annual Conference of the Association of Management, 11(1), 80-85.

Rivlin, A. M. (1973). Measuring performance in education. In M. Moss (Ed.), The measurement of economic and social performance (pp. 411-437). New York: National Bureau of Economic Research.

Roberts, K. H. & Burstein, L. (1980). Issues in aggregation. San Francisco: Jossey-Bass.

Rossmiller, R. (1986). Resource utilization in schools and classrooms: Final report. Madison, WI: Wisconsin Center for Educational Research, University of Wisconsin. (ERIC Document Reproduction Service No. ED 272 490)

Rumbaugh, S. A., Donovan, D. L., Huyser, R. J., & Schooley, D. E. (1972, April). A study of the relationships among district size, expenditure level and composite achievement. Paper presented at the meeting of the American Educational Research Association, Chicago, Illinois. (ERIC Document Reproduction Service No. ED 063 645)

Salmon, R. G., Dawson, C, Lawton, S. B., & Johns, T. L. (1988). Public school finance programs of the United States and Canada. Blacksburg, VA: American Education Finance Association.

Samuelson, P. A., & Nordhaus, W. D. (1989). Economics (13th ed.). New York: McGraw-Hill.

Sebold, F. D., & Dato, W. (1981, January). School funding and student achievement: An empirical analysis. Public Finance Quarterly, 9(1), 91-105.

Spencer, B. D., & Wiley, D. E. (1981). The sense and the nonsense of school effectiveness. Journal of Policy Analysis and Management, 1(1), 43-52.

Stiles, W. D. (1991). The identification of variables that relate to academic success as measured by the Missouri Mastery Achievement Test.

Stinson, T. F. & Krahmer, E. F. (1969). Local school expenditures and educational quality: A correlation analysis. Washington, D.C.: Economic Research Service, Department of Agriculture. (ERIC Document Reproduction Service No. ED 058 384)

Sullivan, H. W. (1974). Supervisor/teacher ratio and achievement in New York schools. (Doctoral dissertation, Hofstra University, 1974). Dissertation Abstracts International, 75-28956.

Summers, A. A. & Wolfe, B. L. (1975). Equality of educational opportunity quantified: A production function approach. Philadelphia: Department of Research, Federal Reserve Bank of Philadelphia.

Testimony of H. J. Walberg. (April 27, 1987). In the Matter of Raymond A. Abbott, et al. vs. Fred G. Burke, et al., 100 N.J. 269, 499 A.2d 376. State of New Jersey Office of Administrative Law, EDU-5581-85. Roseland, NJ: Essex-Union Reporting Services.

Thompson, J. A., & Correa, L. H. (1989, Winter). A study of school and teacher inputs on student achievement outputs. Journal of Education Finance, 14, 390-406.

VanValkenburg, L. E. S. (1990). Relationahip between community wealth and student performance on the Michigan Educational Assessment Program. (Doctoral dissertation, University of Toledo, 1990). Dissertation Abstracts International, 91-04089.

Verstegen, D. A., & Salmon, R. G. (1991, Spring). Assessing fiscal equity in Virginia: Cross-time comparisons. Journal of Education Finance, 16, 417-430.

Verstegen, D. A., & Salmon, R. G. (1989, Fall). The conceptualization & measurement of equity in school finance in Virginia. Journal of Education Finance, 15, 205-228.

Verstegen, D. A., & Salmon, R. G. (1988, Fall). Virginia education finance reform: o8Have excellence and equity been achieved? Journal of Education Finance, 14, 200-220.

Virginia Department of Education. (1992). Outcome accountability project: 1992 Virginia summary report. Richmond: Author.

Walberg, H. J., & Fowler, W. J., Jr. (1987). Expenditure and size efficiencies of public school districts. Educational Researcher, 16(7), 5-13.

Walberg, H. J. (1982). Educational productivity: Theory, evidence, and prospects. Australian Journal of Education, 26(2), 115-122.

Walberg, H. J. (May, 1984). Improving the productivity of America's schools. Educational Leadership, 41(8), 19-27.

Walberg, H. J. (1990). Productivity and equity. In H. P. Baptiste, H. C. Waxman, J. W. de Felix, & J. E. Anderson (Eds.), Leadership, equity, and school effectiveness. Newbury Park, CA: Sage.

Walberg, H. J. (January, 1992). The knowledge base for educational productivity. International Journal of Educational Reform, 1(1), 5-15.

Walberg, H. J., Harnisch, D. L., & Tsai, S. L. (1986). Elementary school mathematics productivity in twelve countries. British Educational Research Journal, 12(3), 237-248.

Walberg, H. J., & Marjoribanks, K. (1976). Family environment and cognitive development: Twelve analytic models. Review of Educational Research, 46(4), 527-550.

Walberg, H. J., & Weinstein, T. (1982). The production of achievement and attitude in high school social studies. Journal of Educational Research, 75(5), 285-292.

Weiskopf, W. A. (1973, Autumn). The image of man in economics. Social Research, 40(3), 547-563.

Weiss, C. H. (1977, Fall). Research for policy's sake: The enlightenment function of social research. Policy Analysis, 3(4), 531-546.

Wendliing, W. & Cohen, J. (1981, Summer). Education resources and student achievement: Good news for schools. Journal of Education Finance, 7, 44-63.

Wirth, A. G. (1988, April). The violation of people at work in schools. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 020 261)

Wolfle, L. M. (1993A). Curvilinearity. Unpublished manuscript. Blacksburg, VA.: Virginia Polytechnic Institute.

Wolfle, L. M. (1993B). Multicollinearity. Unpublished manuscript. Blacksburg, VA.: Virginia Polytechnic Institute.

AUBREY HAMPTON PRICE

Permanent Address:
105 Carol Court
Collinsville, VA 24078

Education:

Virginia Polytechnic Institute and State University
Ph.D. Candidate
Certificate of Advanced Graduate Studies (CAGS), Educational Administration, 1992

University of Virginia
Master of Arts, History, 1972

Virginia Polytechnic Institute and State University
Bachelor of Arts, History, 1968

Academic Achievements:

Doctoral Student, Educational Administration, Virginia Polytechnic Institute and State University, 1992-1993.
Recommended for Doctoral Program (History), Graduate Committee, University of Virginia, 1972.
Graduated with Honors, Virginia Polytechnic Institute and State University, 1968.
Phi Kappa Phi National Honor Fraternity, 1967.

Occupational Experience:

Graduate Assistant, Department of Educational Administration, Virginia Polytechnic Institute and State University, 1992-93
Principal, Fieldale Elementary School, 1988-1992
Principal, Mary Hunter Elementary School, 1986-1988
Principal, Fieldale-Collinsville High School, 1981-1986
Assistant Supervisor of Transportation, Henry County Public Schools, 1978-1981
Teacher and Assistant to the Principal, Richard H. Clark School, 1975-1978
Teacher, Charlottesville Educational Foundation, 1972-1975
Teacher, G. W. Carver High School, 1969-1971

Published Works and Products:

Price, A. H., Strickland, D. C., & Fortune, J. C. (1993). *A critical review of Walberg and Hanushek's contribution to school equity issues.* Forthcoming in The Proceedings, published by the Association of Management.

Fortune, J. C., Strickland, D. C., & Price, A. H. (1993). *Methodological differences in the use of educational productivity function analyses.* Forthcoming in The Proceedings, published by the Association of Management

A Study of Computer Use in the Elementary Schools of Henry County, 1989, Collinsville, Virginia.

A Study in Tire Survival, 1988, Collinsville, Virginia.

Computerized Accounting and Route Database, 1988, Henry County Public Schools.

Computerized Payroll Design, 1987, Transportation Department, Henry County Public Schools.

Price, A. H. (1972). *Three Ambassadors in Berlin, 1933-38: The Observations of Sir Eric Phipps, Andre Francois-Poncet, and William Dodd*, 1972, Charlottesville, Virginia. (Master's Thesis)

Early Background:

Native of Henry County, Virginia
Graduated from Fieldale High School, 1964
Senior Class President, Beta Club President, Editor of School Newspaper, Student Council Officer, Letters in Basketball, Baseball, and Band

Aubrey H. Price