

# METHODOLOGICAL ISSUES IN THE MULTI—LEVEL ANALYSIS OF SCHOOL ENVIRONMENTS

Jean Stockard

---

## ABSTRACT

For a full understanding of student achievement it is essential to use multi-level analyses, which take into account the characteristics of individual students, as well as the nature of their families, classrooms, schools, and communities. Such analyses can be quite complex; and this article reviews, in non-technical terms, relevant methodological issues. Over the last 30 years statistical analyses available for multi-level models have improved greatly. While researchers once used cross-tabulations and, somewhat later, variants of the general linear model, they can now use hierarchical linear models. These techniques allow the exploration of very complex interaction effects, which are often theoretically expected in multi-level models of achievement. Issues of measurement and model specification can also be more difficult with multi-level than with single-level models and need to be carefully considered. It is suggested that a better understanding and use of multi-level models can help bridge the gap between the "input-output" and "process-product" traditions of educational research.

---

Advances in Research and Theories of School Management and Educational Policy,  
Volume 2, pages 217-240.

Copyright © 1993 by JAI Press Inc.

All rights of reproduction in any form reserved.

ISBN: 1-55938-253-8

An overwhelming amount of evidence indicates that students' achievement is influenced by individual characteristics, such as their ability and socioeconomic background. Yet, a great deal of research indicates that classroom, school, and community environments also affect students' learning. For instance, among students with equal measured ability and similar socioeconomic backgrounds, those who are in classrooms and schools with high achievement related norms and more supportive interpersonal environments tend to have high achievement. Similarly, students who live in communities whose citizens support and participate in school activities have higher achievement than students in other communities, even when they have equal individual characteristics. (Stockard & Mayberry, 1992, for an extensive review of this literature.) Thus, most educational researchers today would probably agree that multi-level analyses are essential if we are to fully understand student achievement. Studies of student learning and achievement must take into account not just students' individual characteristics, but also the environments in which they learn.

Analyzing multilevel effects, however, is far from simple. It involves careful attention to the proposed models which are studied, the data used to measure variables in these models, and the techniques used to analyze these data. While there are still many unanswered questions in this area, understanding of the complexities involved has expanded greatly in recent years. Some of these discussions are highly technical and statistical, while others are more accessible to all researchers. Unfortunately, the most recently developed analysis techniques, which are much better suited than earlier methods to handle the various theoretical and substantive complexities, are unfamiliar to many researchers. They are not yet covered in standard statistics textbooks nor included in general statistical packages, and many researchers do not understand how these techniques are related to more familiar methods.

In this paper I review this literature, presenting the material in general terms, with references to the more technical literature for those who are so inclined. I first describe analysis techniques that have been used over the years to examine environmental influences on students' learning. Second, I examine the translation between theory and research design, called the specification of theory; and then explore the issue of measurement, how theory is translated into data. Finally, I briefly describe how multi-level analyses, and especially the most recently developed analysis techniques, can begin to bridge the unfortunate gap between what are commonly called the "input-output" and "process-product" research traditions in studies of student achievement.

My focus is primarily on quantitative analyses, rather than qualitative or ethnographic/field work. This is not meant to denigrate the latter area of research, for such methods are the only way in which to obtain detailed, subjective, rich accounts of how students learn within classrooms, schools, and communities. Moreover the questions and issues I discuss regarding

measurement and specification of models are not unique to quantitative work. Still, most of the literature regarding methodological issues has focused on quantitative analyses, and thus I primarily discuss that literature.

## ANALYSIS TECHNIQUES

The analyses used to examine environmental influences on students' learning developed and changed over the years along with advances in computer technology. The earliest analyses used only aggregated data. For instance, researchers examined the relationship between the average achievement of students within a school or classroom and their average socioeconomic status and the average educational level of their teachers. Unfortunately, such analyses could properly lead only to generalizations about aggregate units, such as schools or classrooms, and could prompt researchers to commit what is known as the "ecological fallacy," a problem discussed in much greater detail below. Recognizing the limitations of analyses that included only aggregated variables, when their true interest was how both individual level and aggregated or environmental variables affected individual students, researchers began to develop multilevel analysis techniques. These could be used to explore hypotheses regarding influences from both individual and aggregated level independent variables on dependent variables measured on the individual level.

The first of these methods were cross-tabulation procedures that could be accomplished with counter-sorters and hand computations. By the 1970s, when computers became more common, most analysts moved to using variations of regression and the general linear model. With the development of very high speed computers in the 1980s iterative techniques building on regression and Bayesian statistics were proposed. Each development has improved researchers' ability to describe the relationships that underlie the influence of environmental variables on students' learning.

### Cross-Tabulation Analyses

In an article published in 1960 and now considered classic, Peter Blau demonstrated how analysts can separate group or contextual influences on individuals from the influences of their own personality or other characteristics. Building on the elaboration techniques popularized by Paul Lazarsfeld and his associates (e.g. Kendall & Lazarsfeld, 1950; Lazarsfeld & Rosenberg, 1955; also see Rosenberg, 1968), Blau demonstrated how the "structural effects of a social value can be isolated by showing that the association between its prevalence in a community or group and certain patterns of conduct is independent of whether an individual holds this value or not" (Blau 1960, p. 180). He suggested that researchers use the characteristics of the individual as

*Table 1.* Illustrations of Blau's Cross-Tabular Analysis of Structural or Contextual Effects

<i>Individual's Plans to Attend College</i>	<i>Individuals' Socioeconomic Status</i>					
	<i>Low School SES Context</i>			<i>High School SES Context</i>		
	<i>Low</i>	<i>High</i>	<i>Total</i>	<i>Low</i>	<i>High</i>	<i>Total</i>
<b>A. Results supporting the Effect of Schools' Socioeconomic Context on Individuals' College Plans</b>						
No	90%	40%	80%	60%	10%	20%
Yes	10%	60%	20%	40%	90%	80%
Total	100%	100%	100%	100%	100%	100%
(n)	400	100	500	100	400	500
<b>B. Results which do not support the Effect of Schools' Socioeconomic Context on Individuals' College Plans:</b>						
No	80%	80%	80%	40%	40%	40%
Yes	20%	20%	20%	60%	60%	60%
Total	100%	100%	100%	100%	100%	100%
(n)	400	100	500	100	400	500

a control variable and then look at the relationship between the contextual variable and the dependent variable within each category of this control variable. If differences persisted then one would be better able to conclude that a group-level or contextual variable had an influence over and above the influence of the individuals' own characteristics. (See also Blau, 1957, as well as Davis, Spaeth, & Husen, 1961 for a slight modification of this technique.)

An example can illustrate Blau's reasoning: Suppose a researcher was interested in studying students' plans to attend college and hypothesized that students with more classmates from higher socioeconomic status (SES) backgrounds would be more likely to aspire to college than students with lower status classmates. This is called a "contextual effect" and has been documented a number of times in the literature (e.g., Alexander & Eckland, 1985; Alexander, McDill, Fennessey & D'Amico, 1979; Mortimore, Sammons, Stoll, Lewis & Ecob, 1988; Stockard & Mayberry, 1992). In testing this hypothesis, it would be important to know that the influence of the socioeconomic context was present whatever the socioeconomic background of the individual was; that is, that the contextual effect was independent of the students' own individual backgrounds. Using the techniques proposed by Blau the researcher would look at data such as that shown in Table 1. (All data are imaginary.)

The results in part A of Table I indicate that a "contextual" effect does occur. Students with higher individual socioeconomic backgrounds are more likely to plan to go to college, no matter what type of school they attend. Yet, all students in the high SES schools, whether they come from lower SES or higher

SES backgrounds, are more likely than those from lower SES schools to plan to attend college. The influence of the environmental or group level variable is independent of the influence of the students' backgrounds. In contrast, the results in part B would indicate that such a contextual effect does not occur. In Part B higher SES students in both types of schools are more likely to plan to attend college than lower SES students and there is no difference in the college plans of students in the two types of schools.

While the example in part A of Table 1 involves individual effects which parallel those on the group level, it is possible that the results on the individual level may be the inverse or opposite of those on the group level. For instance, individual students with higher socioeconomic backgrounds may be more likely to plan to attend college. Yet, students in higher socioeconomic schools might be less likely to plan college attendance, even with similar levels of ability or individual social status, perhaps because of the intense competition they feel within their schools (cf. Davis, 1966; Bachman & O'Malley, 1986; Stockard & Mayberry, 1992). In addition, results on the individual level might be contingent on or vary with the group level value. For instance, participation in a high SES school might positively affect individuals from a lower socioeconomic background, but participation in a lower SES school might not affect students from a high socioeconomic background at all. Following the tradition of elaboration analysis (see Kendall & Lazarsfeld 1950, Rosenberg, 1968), the techniques suggested by Blau could help researchers find such patterns.

The cross-tabulation method is intuitively appealing and was widely used. It also represents a clear improvement over simply looking at aggregated and single-level analyses of student learning. Yet the approach required that variables on the individual level be categorized in some manner. To a large extent any such categorization must be arbitrary in nature and cannot fully represent the true or underlying distribution of a variable. Thus these analyses can produce misleading conclusions about the relative presence of group and individual effects (Boyd & Iversen, 1979; Tannenbaum & Bachman, 1964).

In a sometimes biting and satirical article published in 1970, Robert Hauser (1970) explored what he calls the "contextual fallacy." This occurs "when residual differences among a set of social groups, which remain after the effects of one or more individual attributes have been partialled out, are interpreted in terms of social or psychological mechanisms correlated with group levels of one of the individual attributes" (Hauser, 1970, p. 659). Because people are not randomly assigned to social groups, any variable that can describe these groupings must be associated with a variety of individual level characteristics. For instance, the kind of school students attend is often highly related to their socioeconomic background and race and ethnicity. In a contextual analysis, such as that shown above, a researcher generally controls for one and usually not more than 2 or 3 individual level variables while looking for the influence

of a group level variable. In so doing the researcher implicitly assumes that the group-level variable is only related to these selected individual level variables and no others. If, in fact, the grouping variable is associated with other individual level variables not included in the analysis (the "residual differences" Hauser refers to), conclusions about the existence of a contextual effect are likely to be faulty.

Hauser's critique hinges on the problem of properly specifying the range of variables that may influence a dependent variable, a problem that can appear with any kind of analytic technique and which we discuss more fully below. Yet the problems associated with grouping of variables are especially noticeable with cross-tabulations and Hauser recommended that researchers interested in contextual and structural effects begin to use techniques associated with the general linear model rather than cross-tabulations.

### General Linear-Model Techniques

By the 1970s computers had largely replaced counter sorters and were widely available to researchers. While analyses belonging to what is called the general linear model (Cohen, 1968; Gordon, 1968), such as multiple regression and complex analyses of variance and covariance, are extremely difficult to do by hand, the computer made it possible to conduct such analyses relatively easily with large amounts of data. Within a few years most researchers agreed with Hauser that such techniques were much more suitable to the analysis of multi-level data than cross-tabulations, even though the cross-tabulation techniques and analysis of covariance, a variation of the general linear model, are essentially equivalent under certain circumstances (Alwin 1976). The techniques within the general linear model allow a researcher to examine the relationship between a dependent variable and a number of independent variables, which can be on varying levels of analysis.

The elements of the general linear model are familiar to most researchers. The simplest form is the bivariate regression equation:

$$Y_i = a + b_{yx} X_i + e_i, \quad (1)$$

where  $Y_i$  is the score or measure on the dependent variable for an individual  $i$ ;  $a$  represents the regression intercept or the predicted value of  $Y$  when  $X=0$ ;  $b_{yx}$  represents the slope of the regression line, or the predicted change in the dependent variable  $Y$  that would occur with a unit change in the independent variable  $X$ ; and  $e_i$  represents the error term for individual  $i$ , or the difference between the actual value of  $Y_i$  and the value predicted by the regression equation.

This basic equation may be expanded in a variety of ways. When several intervally measured variables are used as predictors, the equation is simply known as a multiple regression equation. When categoric variables are used as predictors in the form of dummy variables and the equation also includes

interaction terms for these predictors, the model is that of analysis of variance. When both continuous and categoric predictor variables as well as their interaction terms are used, the model becomes that of analysis of covariance.

Because of its flexibility this model can easily be expanded to encompass analyses of many of the relationships that are inherent in descriptions of environmental influences on learning. For instance, many studies examine how student achievement is influenced by variables related to individual students, such as measures of their ability, socioeconomic status, and previous achievement; and by variables related to the groups in which these students learn, such as the average socioeconomic status of the class or the school and other characteristics of the classroom, school, or community. A model could be developed that includes each of these variables.

As a simple example, suppose that a researcher wanted to predict  $Y$ , a measure of student achievement, from students' measured ability ( $X_1$ ), their socioeconomic status ( $X_2$ ), their teachers' verbal ability ( $X_3$ ), and the average socioeconomic status of their school ( $X_4$ ). The resulting prediction equation would be

$$Y = a + b_{y1} X_1 + b_{y2} X_2 + b_{y3} X_3 + b_{y4} X_4. \quad (2)$$

Estimating the regression coefficients (the  $b$ 's) through a technique called ordinary least squares yields results that describe the extent to which changes in each of the independent variables (ability, socioeconomic status, teachers' verbal ability, and the average SES of the school) influence changes in the dependent variable (students' achievement), independent, or net, of the influence of each of the other independent variables.

A wide variety of analyses are possible within this general model, each dependent on the hypotheses which a researcher wants to test (see discussion below on specification and Boyd & Iversen 1979; Burstein, 1980, 1981; de Graaf, 1984; Firebaugh, 1980; Stipak, 1980; Stipak & Hensler, 1982). More complex extensions of this general technique, such as structural equation, dynamic change, and multiple time series models, allow researchers to include feedback loops within the analysis and to examine longitudinal data dealing with such issues as changes in achievement over time (see Rogosa 1980). Although relatively rare still in the literature on student learning, it is also possible to extend the regression model to the analysis of categoric dependent variables with the use of discriminant function analysis and varieties of log-linear modeling (see Goodman, 1978). Structural equation models and the popular LISREL program, also allow researchers to explore hypotheses concerning the extent to which measurement error affects the results and to use multiple indicators of measures (see Hayduk, 1987). These models are extremely flexible and probably among the most useful of the analysis techniques to arise from the general linear models.

By the 1980s most researchers agreed that if a researcher was interested in

studying students, rather than using schools or classrooms as the unit of analysis, multilevel analysis using general linear model techniques was preferable to cross-tabulation methods (e.g. Hopkins, 1982; Lincoln & Zeitz, 1980). The analyses based on the general linear model have become widely accepted and are commonly included in statistics textbooks and discussed in methodology classes. Yet, because they are not specifically designed to deal with multilevel analyses, they cannot deal with all of the concerns and complexities involved in understanding environmental influences on individuals.

### Hierarchical Linear Models

The most recent analytic techniques suggested for use with multilevel data are called by names such as Hierarchical Linear Models, Random Coefficient Models, and Empirical Bayes Techniques. These are potentially very powerful tools that only became available for widespread use with the development of very high speed computers in the 1980s. To date the techniques have been used by only a few educational researchers and, unlike the general linear and structural equation models, are not available in commonly used statistical packages nor reviewed in statistics textbooks.

Hierarchical linear models can deal with a problem that has not been solved with the general linear model techniques. When researchers use these traditional analyses to examine how both individuals' characteristics and school or classroom policies or procedures affect behaviors of individual students they must assume that the influence of the individual level independent variables is consistent from one school or classroom to another. For instance, a researcher may examine the influence of socioeconomic status (SES) on students' achievement in a large number of both public and private schools, but the general linear model techniques do not allow the researcher to explore the possibility (if not probability) that the influence of SES on achievement varies from one school to another and, more importantly, to develop models that can explain why the influence of SES varies from one school to another.

The hierarchical linear models help fill this gap. They use the basic outlines of the general linear model, predicting a dependent variable from a series of independent variables that reflect not just the level on which the dependent variable is measured but also those at higher or more aggregated levels of analysis. The researcher first estimates the model in essentially the same way one would with the general linear model. Then, however, the researcher proceeds to use the parameters (the regression coefficients) that were estimated at this first level of analysis (usually the individual level in studies of educational effects) as dependent variables and the variables from higher levels of analysis as independent variables. The results of this second stage of the analysis allow the researcher to see how school and/or classroom related variables affect the way in which individual level variables influence achievement. That is, the model



allows the researcher to directly estimate how group level variables, such as characteristics of the school or classroom, influence the way in which individual level variables, such as individuals' SES, affect achievement.

There are a variety of ways of calculating the parameters in these models, all of which involve a series of iterations, or repeated calculations, to develop estimates that most accurately reflect the data. These estimates are developed through the use of Bayesian statistics and can be calculated with the aid of computer programs developed by various authors in the field. (See Bryk & Raudenbush, 1989b; Burstein, Kim & Delandshere, 1989; Mason, Wong, & Entwisle, 1983; Raudenbush & Bryk, 1986, 1988, 1989 for extensive discussions of these methods.) Because the techniques are not yet widely used, relatively few examples of their utility for studies of environmental influences on achievement are in the literature. However, the methods appear to be very well suited for the theoretical issues that face researchers interested in this area. They can handle not just estimations of effects at different levels of analysis, but also changes in effects over time (see Bryk & Raudenbush, 1989b), categorical dependent variables (Wong & Mason, 1985), and, to some extent, estimates of measurement error (DiPrete & Grusky, 1990).

No matter what type of analysis technique a researcher chooses to use, adequate hypotheses and measures are a necessity. No analysis, no matter how sophisticated, can compensate for poor specification of theories or poor measurement, and I turn now to these issues.

## **SPECIFICATION OF THEORETICAL MODELS**

Specification refers to the precise delineation of theoretical models in a testable format. It is especially problematic when dealing with environmental influences precisely because researchers are usually concerned with more than one level of analysis. Sometimes specification errors result because researchers have used aggregated data rather than data that more closely matches their theories, and we first discuss issues related to the use of aggregate data. We then move to a more general discussion, examining common specification errors and how these may affect analyses of environmental influences.

### *Issues in Aggregation*

Aggregated data have been used in a variety of ways, some more appropriate than others. Below I discuss reasons researchers use aggregated data and criticisms of their use.

#### *Why Aggregated Data are Used*

Aggregated data have often been employed because there were no alternatives. Researchers, actually interested in studying individual level variables, such as student achievement, have been forced by the nature of the available data to examine variables measured at higher levels of analysis. Thus, instead of looking at the achievement of individual students, they have examined average student achievement in classrooms or schools. If the researcher's theoretical interest is actually at the level of the individual, using data from a higher level can, as we show below, introduce biases and produce misleading results.

It is, however, possible that researchers might choose to examine data at an aggregated level because their theoretical interests lie in explaining processes that occur at that level. Those who advocate using the individual as the unit of analysis focus on the fact that students learn and react to school environments as individuals and that even environmental influences are often only an aggregation of individual effects (Wittrock & Wiley, 1970). Those who advocate using classrooms as the unit of analysis assert that this is appropriate when evaluating the effect of classroom instruction because all students are simultaneously exposed to the treatment variable, or instruction, and it is important to examine the effect of this treatment at the level where it occurs (e.g., Wiley, 1970). Those who advocate using schools or districts as the unit of analysis stress the importance of policy decisions that are based at the school or district level (e.g., Bidwell & Kasarda, 1975, 1976). As long as the theoretical arguments are centered around the level of aggregation at which the data are gathered and analyzed, the work may resist criticism.

Scholars might also choose to analyze experimental data at an aggregated level because they are concerned about the independence of the units that are studied. Experimental treatments or variables are often administered to students within classrooms. Because students received experimental treatments within groups, it was once argued that it is appropriate to analyze the effect of the experimental variable using the groups, generally classrooms, as the unit of analysis (Lindquist, 1940). More recent examinations of this area, however, suggest that the problem has been overstated and that such experimental data may be more appropriately analyzed by using a multi-level analysis such as those described above (Hopkins, 1982).

### *Criticisms of Aggregated Data*

The first widely read and influential criticism of the use of aggregated data was developed by William Robinson (1950) in his discussion of the "ecological fallacy." He described the problems that can result from using data on an aggregated level of analysis to generalize to a lower level of analysis, for instance, using correlations based on school level data to generalize to individuals within those schools. Robinson demonstrated that correlations based on grouped data

can often overestimate the correlations obtained when individual data are used and suggested that great care be taken in drawing conclusions from studies using aggregated data. Using aggregated data rather than individual level data essentially destroys information about the individual involved and reduces the variability of the variables. It can also mask nonlinear effects that might be present in the data and assumes that all students within a group are essentially similar on the range of characteristics considered. In general, studies that rely only on aggregated data can produce inconsistent estimates of effects. This problem can appear with both simple and multiple regression models, non-recursive models, and longitudinal models (Burstein, 1980).

Later analyses of the "ecological fallacy" have expanded upon Robinson's conclusions by demonstrating that aggregation bias, or the bias that comes from inferring across levels of analysis, can be understood by analyzing the way in which the grouping of cases relates to the distribution of the variables involved in the analysis. Grouped-level and individual level estimates of effects are only different from each other when the way in which the cases are grouped is related to scores on the independent and/or dependent variables. Unfortunately for researchers, this often happens in studies in education. For instance, classroom groupings are often based on children's prior achievement; the school to which children are assigned is often related, through patterns of neighborhood segregation, to their socioeconomic status and racial and ethnic background. The fact that these individual characteristics are related to both the groups to which children are assigned as well as to individual level variables of interest, such as achievement, means that aggregation bias is likely to occur. Conclusions derived from analyses on an aggregated level cannot be extended to an individual level.

Fortunately, the extent of this aggregation bias may be estimated with a variable which measures group membership. This may be done either directly, through entering the grouping variable into the analysis (see Burstein, 1978; Hannan & Burstein, 1974), so that estimates of effects include an estimate of the effect of grouping itself, or more indirectly, by including the group mean on the independent variable within the analysis (Firebaugh, 1978). Both techniques yield similar results (Burstein, 1980, 1981).

On a more general level, a number of authors have suggested that estimates of effects differ across levels of aggregation because the theoretical models have been misspecified. In other words, aggregation bias is simply specification bias (Hanushek, Jackson & Kain, 1974; Irwin & Lichtman, 1976; Langbein & Lichtman, 1978). If all of the individual level variables that actually influenced the dependent variable were included in a model, any bias introduced by grouping should be nil. Certainly such an argument makes logical sense, and it has been suggested that the methods for checking for aggregation bias are in reality a search for specification errors.

### Typical Specification Errors

Even though researchers now increasingly realize the importance of specifying multi-level models in analyzing educational outcomes, many potential methodological problems related to specification remain. The most often cited problem is that of omitting relevant variables from the analysis. We first examine this area and then discuss theoretical explanations of group effects and other problems.

#### *Omission of Relevant Variables*

Omitting relevant variables is a very serious problem because it is likely to directly bias estimates of the effect of the variables of interest. The statistical explanation of why these biased estimates occur can best be understood by considering analyses that use a regression model such as that described in equations (1) and (2) above. In an equation such as (1),  $e_i$  represents the error term, or the extent to which the predicted value of  $Y$  deviates from the actual value of the dependent variable  $Y_i$  for that case. Essentially,  $e_i$  represents all possible influences on the dependent variable other than those included in the model.

The regression model (and actually all techniques of data analysis) assumes that the error term in the equation is uncorrelated with the independent variables. That is, any possible influences on the dependent variable other than those represented by the independent variables are totally unassociated with the independent variables already in the equation. The estimates of the regression coefficients are correct only if this assumption is true. If the error term is associated with any of the predictor variables, the estimates in the regression equation are said to be biased and the equation is said to be misspecified.

The possibility of such misspecification is unfortunately quite large when survey data are used to analyze environmental influences on learning. While strictly constructed experimental designs can control for the possibility of correlated error terms, such designs are usually impractical in educational research. Most large-scale studies of student learning, apart from very small laboratory settings, involve the use of survey-based data and intact groups. When researchers wish to estimate the effect of environmental variables, such as the average ability level or the average socioeconomic status of a group, on students' achievement, any estimate of that effect will be biased unless the specified model includes all other variables that are conceivably associated with both the dependent variable and the hypothesized independent variable. For instance, characteristics of teachers in a district may be associated with characteristics of the community and students, with higher status communities and students of higher ability more likely to attract teachers with more extensive

qualifications. Any estimate of the effect of teachers' characteristics on students' achievement would then need to come from a model that incorporated not just variables related to the students, but also those related to the school and the community.

This problem can occur no matter what level of analysis is involved. But, when researchers are dealing with data from several levels of analysis the probability of specification errors may be increased. For instance, a researcher might suggest that the achievement of individual students could be accounted for by the educational level of teachers and the quality of a school's library. Yet, all three of these variables might well be associated with students' ability, attitude toward schooling, and individual socioeconomic status. If the researcher did not also include these individual level variables in the model the estimate of the effect of teachers' education and the school library on achievement would be biased and incorrect. (For more extensive discussions of the statistical problems involved in this area see Burstein, 1980, 1981; for a fascinating debate regarding the effect of specification errors on an analysis see Bidwell and Kasarda, 1975, 1976, Hannan, et al., 1976; Alexander & Griffin, 1976).

### *Explaining Group Effects*

This statistical explanation of why omitting variables from a model of environmental effects can produce biased results holds across all discussions of environmental influences. Yet, theoretical distinctions can be made regarding how group effects occur. That is, group effects may result simply from how people are assigned to groups (selection effects) or they may reflect group processes or properties.

The input-output tradition of research has tended to focus on identifying selection effects. These selection effects reflect the fact that students are nonrandomly assigned to classrooms, schools, and school districts. Race, ethnicity, socioeconomic status and ability can all affect these assignments. To the extent that these individual-level variables are associated with both the independent and dependent variables in a model, the results may be misleading. In other words, if the assignment of students to schools and classrooms is related to background variables that also influence achievement, any differences in achievement between classrooms and schools may reflect these background differences rather than any effect of the schools or classrooms themselves. A properly specified model, which allows for the influence of both the background variables and the grouping variables, can indicate the independent effect of both types of variables on students' learning.

Only when the effect of grouped level variables on learning persists after the influence of selection related variables is controlled is it possible to suggest that some group process is influencing student achievement. The process-

product tradition of educational research has focused on trying to understand more how group processes and interactions influence educational outcomes. Yet, as I discuss more below, this tradition has generally not simultaneously tried to rule out the possibility that selection effects might be affecting the results.

Most discussions of specification errors tend to focus on the problems of overestimating the effects of group-level variables (e.g., Hannan, 1971; Robinson, 1950), focusing on the problems inherent in biases related to the differential selection of students into learning groups. A few scholars, however, have altered this discussion by noting that when we control for individual level variables we may tend to underestimate the actual effect of schools and other group level effects (e.g., Bidwell & Kasarda, 1980, p. 403; Centra & Potter, 1980, p. 277). The authors are making a theoretical point and choosing to focus on the ways in which groupings affect the processes within schools. Individual level variables, such as students' socioeconomic background, as well as group level variables, such as the schools to which students are assigned, influence students' achievement. But these variables also influence the process of schooling itself. Multi-level and cross-level analyses of students' learning must take into account not just how students' backgrounds affect their individual rates of learning, but also how groupings of students affect the process of teaching and learning which occurs within classrooms and schools. In the final section of this paper I suggest that the most recently developed statistical techniques may help researchers actually carry out analyses such as these.

### *Other Problems*

While the omission of important variables is undoubtedly the most common and potentially serious specification error in examining cross-level data, a number of other problems may also occur. First, the time order of variables and the specification of cause and effect may be faulty (cf. Eckart & Durand, 1985). For instance, a researcher may suggest that more highly qualified teachers tend to enhance students' learning, while in fact, more capable, harder working, and higher achieving students tend to attract more qualified teachers. The direction of causality may be exactly opposite to that hypothesized.

Related to this concern is a common failure, often attributable to limitations of available data, to include time dimensions in hypothesized models. The possibility of simultaneous or reciprocal casual effects as well as longitudinal influences should be considered. If the characteristics of students and teachers within schools and districts could be tracked over time, it would be possible to more clearly understand the reciprocal relationship between these two variables. It is possible that studies that take the cumulative nature of schooling into account actually indicate a greater degree of influence of environmental influence (e.g., Heynes, 1978; see also Stockard and Mayberry, 1992). Because

students learn over time it is important that models of student learning incorporate this longitudinal aspect (cf. Bidwell & Kasarda, 1980; Centra & Potter, 1980, p. 277; Eckart & Durand, 1985, p. 3-6; Stipak, 1980, p. 47).

Most analyses of environmental effects also specify relationships as linear in nature. In fact, however, they may not approximate this form at all, and constraining the analysis in this manner may falsely estimate the actual effect. The literature on class size illustrates this possibility, where most of the influence of decreasing class size on achievement occurs only when classes are smaller than approximately 15 students (Glass, Cahen, Smith, & Filby, 1982). Only studies that take this curvilinear effect into account can demonstrate this relationship (cf. Centra & Potter, 1980, p. 277).

## MEASUREMENT ISSUES

As noted earlier, most researchers now agree that it is usually best to use multi-level analyses when examining environmental influences on student learning. In general, the literature on measurement suggests that one should usually measure the dependent variable of student learning on the individual level and the explanatory variables at levels that represent their theoretical influence (cf. Bidwell & Kasarda, 1980; Burstein, 1980, 1981; Stipak & Hensler, 1982). However, if researchers are to make accurate conclusions about environmental influences, they must not only correctly specify the theoretical models to be examined and analyze them in an appropriate manner, they also must correctly measure the variables that are involved. Measurement rules apply to all research endeavors but are, if possible, even more crucial when examining environmental influences and multi-level models. Errors and inadequacies at one level simply become compounded when extended to another level, especially if composite measures are used. Below we examine issues related both to adequacy of measures and measurement error.

### Adequate Measures

When one considers the adequacy of measures it is important to consider both how the measures are related to the theory one hopes to test and how the measures are constructed.

### *Matching Theory and Measures*

Measures must accurately reflect the theory which is to be tested. Social scientists often use measures that only begin to tap the full complexity of theoretical concepts. Yet, if they are to adequately test their theories it is important that researchers attempt to operationalize concepts in ways that reflect their theories as closely as possible. For instance, a researcher might

hypothesize that the socioeconomic composition of a classroom influences the expectations which students receive in their interactions with each other and with teachers. To adequately measure the concepts in the theory, it would be important to measure not just the socioeconomic composition of the classroom, but also the expectations which students and teachers hold regarding achievement. The resulting model would much more accurately reflect the researcher's actual theory and thus would provide a more accurate test.

It is also important to realize that the conceptual meaning of variables may differ from one level of analysis to another. For instance, on the individual level of analysis a measure of socioeconomic status (SES) reflects the individual resources that students bring to the learning situation. On the level of the classroom an aggregated measure of socioeconomic status may reflect the expectations that teachers have for the students, the type of curriculum assigned, the nature of peer expectations, and implicit messages regarding success. At the level of the school and district an aggregated measure of SES may reflect the financial resources of a community.

In addition, studies of the effects of environment on students usually specify that measures of the central tendency of these groups indicate their characteristics, focusing on the average socioeconomic status or average ability of students in a classroom or school. Yet, many times it may be aspects of the groups other than central tendency that affect the process of schooling. Groups that are very homogeneous may require different teaching methods and produce different environments than groups that are very heterogeneous. Analyses that take into account this variability would benefit from considering measures of variance rather than measures of central tendency. Similarly, groups that deviate a great deal around a rather low mean would probably have substantially different environments than those that deviate a smaller amount around a larger mean. Here, a measure of inequality, such as a coefficient of variation, might be an appropriate indicator (cf. Burstein, 1981, p. 214-215; Stipak & Hensley, 1982, p. 155-157). The hierarchical linear model techniques discussed above seem especially well suited to analyzing dispersion of a dependent variable (see Bryk & Raudenbush, 1989 a & b; Raudenbush, 1988; Raudenbush & Bryk, 1987).

### *Standard Rules of Measurement*

While all researchers, whether dealing with one level of analysis or several, must endeavor to follow standard rules of measurement, some are especially important to those studying environmental influences on student achievement. Two of the most crucial involve the variation of the measures and categorization. Estimates of the effects of environments upon individuals may be underestimated if the groups or environments which are studied are not sufficiently variable or different from each other. For instance, estimates of



the effects of resources and facilities on student learning may be underestimated in the current literature because most studies include schools with only a small amount of variation on these variables. Regulations regarding certification of teachers and accreditation of schools ensure that teachers have fairly similar formal qualifications and schools have relatively similar facilities. Studies in countries other than the United States that have incorporated more variation on these variables typically find much larger effects than studies in the United States (Stockard & Mayberry, 1992). Similarly, contradictory results in studies of the effect of class size may be directly traced to the lack of sufficient variation in the independent variable (Glass, et al., 1982; Stockard & Mayberry, 1992).

Another standard rule of measurement (and analysis) is that researchers should preserve the original metric of measures and not needlessly categorize variables. If a variable is conceptually continuous in nature the measure of that variable should reflect, as much as possible, that metric. As we noted above, the rather arbitrary categorization of variables for cross-tabulation analyses of contextual effects could produce faulty results (see Alwin, 1976; Hauser, 1970). Today, when cross-tabulation techniques are less widely used, researchers may use categoric measures rather than interval or ratio level measures simply as a matter of convenience. If we are to have accurate tests of our theories, however, it is important that researchers strive to measure variables in ways which accurately reflect the underlying metric.

#### Measurement Error

As with the issue of adequacy, the issue of measurement error affects all social research. With multi-level analyses, some aspects are even more difficult.

Research on student achievement is often conducted within schools and intact classrooms. Even when all students in a school or a classroom are not studied, cluster sampling is often used to minimize research costs. These procedures almost inevitably result in clusters or units from which students are sampled, whether they be schools or classrooms, which are more homogeneous than the total population. Because of the way students are assigned to classrooms and schools, they tend to be more similar to each other than to the total population.

In addition, in statistical language, the observations are dependent upon each other or errors are correlated. Students within classrooms have many experiences in common: the same teachers, the same peers, usually the same community background and often similar social class and racial-ethnic backgrounds. They interact with each other on a daily basis, and thus the observations of their behavior cannot be considered independent. The variance of the error term may also differ across groups. For some groups an independent variable, such as socioeconomic status, may be a strong predictor of a dependent variable such as achievement, while for others it may be

somewhat weaker. As we noted above, when either the error terms are correlated or the variance of the error term differs across groups, the value of the regression coefficients will be inaccurate (see Burstein 1981, pp. 192-3). A technique such as hierarchical linear modeling should be used to assess these differential effects.

Special estimation problems arise when contextual values or aggregated variables estimated from sample data on individuals are used in conjunction with the data from individuals. For instance, a researcher interested in how socioeconomic status is related to achievement may estimate the socioeconomic status context of a school (symbolized here by  $S$ ) by taking the average of the measures of socioeconomic status for each of the students ( $s$ ) through the simple procedure shown in equation (3).

$$\sum s/n = S \quad (3)$$

The researcher may then examine how both this aggregated measure of socioeconomic context ( $S$ ) and the measure of socioeconomic status from the individual level ( $s$ ) affect students' achievement ( $A$ ), as in equation (4).

$$A = a + b_{AS} S + b_{As} s \quad (4)$$

In this situation the sampling error for the independent variable measured on the aggregated level ( $S$ ) is correlated with the sampling error for the corresponding variable on the individual level ( $s$ ) because both variables come from the same underlying measurement. This violates the basic assumption of the linear model involving independence of the error term and can lead to inaccurate estimates. The fewer the number of cases used to estimate a contextual variable, the worse this problem becomes.

One possible solution to this problem is to estimate the value of the contextual variable for each individual by omitting the individual's own value. That is, the measure of the school's socioeconomic context would be estimated from an average of the socioeconomic status of cases other than that individual and would be different for each individual. Although it is not always possible to do so, a far better solution would simply be to gather independent estimates of the contextual variable itself. That is, the researcher could try to assess more directly the aggregate variable through the use of other observations and indicators. This could also result in more valid measures of the aggregate variable.

Researchers have traditionally used estimates of reliability and validity to estimate the extent of error within their measures. While reliability and validity of measures should always be of concern to researchers, such issues have often not been confronted by those dealing with environmental effects. Validity of variables may be assessed in a variety of ways, with, as discussed immediately above, attempts to develop several independent indicators of environmental influences being perhaps the most important (Zeller & Carmines, 1980).

Reliability of environmental variables which are not composites of those from lower levels may be assessed through traditional means. It is also possible to estimate the reliability of composite variables (those aggregated from scores on a lower level) from the aggregate level data through methods developed by O'Brien (1984, 1990, 1991).

## **BRIDGING GAPS IN EDUCATIONAL RESEARCH**

Two major areas of educational research have examined environmental influences on student achievement. Some analysts use what has been called an "input-output" approach, focusing on the ways in which student, classroom, and school characteristics or resources (the "inputs") influence student achievement (the "outputs"). Others use a "process-product" orientation, which focuses on how the actual behaviors of teachers and students (the "process" within a classroom) influence student achievement (the "product"). Much useful research has resulted from both orientations. The process-product approach seems especially well suited for exploring the actual process by which students learn, the way in which school resources and environmental influences are translated into individual students' learning. The input-output approach seems especially well suited for analyzing data from larger surveys and estimating and summarizing general patterns of influences of environmental effects.

While most researchers would probably agree that the variables included in both the input-output and the process-product approaches are important influences on achievement, there has been little movement toward merging these two traditions, largely perhaps because their methodological approaches have been fairly distinct. Input-output research tends to utilize large-scale survey data and/or school and district based records of achievement and resources, as well as relatively sophisticated statistical analyses. Many of the methodological advances described in this paper grew out of the input-output tradition. The process-product approach generally relies on classroom and student observations and is less likely to use large samples and sophisticated statistical techniques.

Theoretically, however, it seems clear that a full understanding of environmental influences on student achievement requires a merger of these traditions. In a recent book, I present a conceptual framework that shows how both the general characteristics of students' environments (the "inputs") and the way in which these characteristics are translated and understood by teachers and students (the "process") influence students' learning (the "output" or "product"). That is, both the "social order" of schools, classrooms, and communities, and the "social actions" of students, teachers, administrators, and parents work together to produce higher achievement (Stockard & Mayberry,

1992). We need to understand these multilevel relationships if we are to continue to promote student achievement.

I believe that the current challenge for educational researchers is to expand upon and examine multilevel conceptual frameworks such as this through appropriate specification of models, accurate measurement of the variables involved, and suitable analysis techniques. Multilevel and integrated conceptual frameworks need to be translated into clearly specified and researchable models of achievement. Researchers need to expand the models of learning which they now have to include variables related to both social order and social action and to be careful to ensure that relevant variables are included. Researchers need to understand that students are embedded within classrooms, schools, and communities and that all of these influence the ways in which they learn.

Multiple measures appropriate for a variety of levels of analysis need to be tested for both validity and reliability. It is especially important that these variables reflect the theoretical constructs specified in the model, that is that they are valid. With multilevel models it is especially important that these measures accurately reflect the construct at the hypothesized level. It is also essential that the measures be reliable—that they can produce consistent estimates of characteristics and behavior when used by different researchers and in different settings.

Finally, we need to use analysis techniques that can incorporate both the analysis of classroom process and the analysis of environmental inputs or structures. I believe that the most recently developed hierarchical models can help to bridge this analytic gap. They allow a researcher to examine how classroom processes, the specific behaviors of teachers and students, influence learning and how larger school environments, including characteristics of classrooms, schools and communities, affect the way in which these influences take place. Thus they can help researchers analyze models that incorporate notions of both how the social actions within schools and the social order of schools, classrooms and communities affect student learning.

Testing such models, however, will only be possible if, and when, data sets are developed which incorporate both types of data. This will probably require researchers to show more communication and flexibility than is sometimes apparent. We need data sets that incorporate the beneficial aspects of both the input-output and process-product traditions, data sets that include broad-based samples and data from a variety of environmental sources as well as in-depth analyses of classroom and school processes.

## SUMMARY

Educational researchers generally agree that students' learning is influenced not only by their individual characteristics but also by the environments in

which they live. Yet, analyzing these multilevel influences is far from simple. Techniques for analyzing multilevel data have become much more advanced in the last 30 years, from the early cross-tabulation techniques, through the various applications of the general linear model to the most recent developments of hierarchical linear models, which allow researchers to examine much more complex interaction effects. Yet, no matter how complex an analysis technique is, it is useless without a properly specified theoretical model and adequate measures. While researchers certainly continue to try to develop better measures and to use care in specifying models, we suspect that these two areas are still in need of a great deal of work. If further advances are to be made in understanding environmental influences on education, it seems extremely important that researchers develop better measures, carefully specify their theories, and work toward developing data sets and using analysis techniques that can incorporate the essential aspects of both the input-output and process-product research traditions.

## ACKNOWLEDGMENTS

I would like to thank Samuel Bacharach, Sally Bowman, Maralee Mayberry, Robert O'Brien, and Joe Stone for helpful comments on earlier drafts of material related to this article. Any errors and all opinions are, of course, my own responsibility.

## REFERENCES

- Alexander, K. L., & Eckland, B. K. (1975). Contextual effects in the high school attainment process. *American Sociological Review*, *40*, 402-416.
- Alexander, K. L., & Griffin, L. J. (1976). School district effects on academic achievement: A reconsideration. *American Sociological Review*, *41*, 144-152.
- Alexander, K., McDill, E., Fennessey, J., & D'Amico, R. (1979). School SES influences—composition or context? *Sociology of Education*, *52*, 222-237.
- Alwin, D. F. (1976). Assessing school effects: Some Identities. *Sociology of Education*, *49*, 294-303.
- Bachman, J. G. & O'Malley, P. M. (1986). Self-concepts, self-esteem, and education experience: The frog-pond revisited (Again). *Journal of Personality and Social Psychology*, *50*, 35-46.
- Bidwell, C. E. & Kasarda, J. D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education*, *88*, 401-430.
- Bidwell, C. E. & Kasarda, J. D. (1976). Reply to Hannan, Freeman, and Meyer, and Alexander and Griffin. *American Sociological Review*, *41* 152-160.
- Bidwell, C. E., & Kasarda, J. D. (1975). School district organization and student achievement. *American Sociological Review*, *40*, 55-70.
- Blau, P. M. (1960). Structural effects. *American Sociological Review*, *25*, 178-193
- Blau, P. M. (1957). Formal organization: Dimensions of analysis. *American Journal of Sociology*, *63*, 58-69.

- Boyd, L. H., Jr., & Iversen, G. R. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, California: Wadsworth.
- Bryk, A. S. & Raudenbush, S. W. (1989b). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 396-404). San Diego: Academic Press.
- Bryk, A. S. & Raudenbush, S. W. (1989a). Heterogeneity of Variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, 396-404.
- Burstein, L. (1981). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, 8, 158-233.
- Burstein, L. (1980). The role of levels of analysis in the specification of educational effects. In R. Dreeben & J. A. Thomas (Eds), *The analysis of educational productivity, Volume 1: Issues in microanalysis* (pp. 119-190). Cambridge, Massachusetts: Ballinger.
- Burstein, L. (1978). Assessing differences between grouped and individual-level regression coefficients: Alternative approaches. *Sociological methods and Research*, 7, 5-28.
- Burstein, L., Kim, K., & Delandshere, G. (1989). Multilevel investigations of systematically varying slopes: Issues, alternatives, and consequences. In R. D. Bock (Ed), *Multilevel analysis of educational data* (pp. 233-276). San Diego: Academic Press.
- Centra, J. A. & Potter, D. A. (1980). School and teacher effects: An interrelational model. *Review of Educational Research*, 50, 273-291.
- Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70, 426-443.
- Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology*, 72, 17-31.
- Davis, J. A., Spaeth, J. L., & Husen, C. (1961). A technique for analyzing the effects of group composition. *American Sociological Review* 26, 215-225.
- deGraaf, C. (1984). Multi-level analysis in educational research. In H. Oosthoek & P. Van Den Eeden (Eds), *Education from the multi-level perspective: Models, methodology and empirical findings* (pp.45-70). New York: Gordon and Breach.
- DiPrete, T. A., & Grusky, D. B. (1990). The multilevel analysis of trends with repeated cross-sectional data. In C. C. Clogg (Ed), *Sociological Methodology 1990* (pp. 337-368). Washington, D.C.: American Sociological Association.
- Eckart, D. R., & Durand, R. (1985). Contextual variables and policy arguments: A critique. *The Social Science Journal*, 22, 1-14.
- Firebaugh, G. (1980). Assessing group effects: A comparison of two methods. In E. F. Borgatta, & D. J. Jackson (Eds), *Aggregate data: Analysis and interpretation* (pp. 13-24). Beverly Hills: Sage Publication.
- Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, 43, 557-572.
- Glass, G. V., Cahen, L. S. Smith, M. L., & Filby, N. N. (1982). *School class size: Research and policy*. Beverly Hills, CA: Sage Publications.
- Goodman, L. A. (1978). *Analyzing qualitative/categorical data: Log-linear models and latent-structure analysis*. Cambridge, Massachusetts: Abt Books.
- Gordon, R. (1968). Issues in multiple regression. *American Journal of Sociology*, 73, 592-616.
- Hannan, M. T. (1971). *Aggregation and disaggregation in sociology*. Lexington, Mass.: Heath-Lexington.
- Hannan, M. T., & Burstein, L. (1974). Estimation from grouped observations. *American Sociological Review*, 39, 374-392.
- Hannan, M. T., Freeman, J. H., & Meyer, J. W. (1976). Specification of model for organizational effectiveness. *American Sociological Review*, 41, 136-143.
- Hannushek, E. A., Jackson, J. & Kain, J. (1974). Model specification, use of aggregate data, and the ecological correlation fallacy. *Political Methodology*, 1, 89-107.

- Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.
- Heyns, B. (1978). *Summer Learning*. New York: Academic Press.
- Hopkins, K. D. (1982). The unit of analysis: Group mean versus individual observations. *American Educational Research Journal*, 19, 5-18.
- Irwin, L., & Lichtman, A. J. (1976). Across the great divide: Inferring individual level behavior from aggregate data. *Political Methodology*, 3, 411-439.
- Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds), *Continuities in Social Research* (pp. 187-195). Glencoe, Ill.: Free Press.
- Langbein, L. I., & Lichtman, A. J. (1978). *Ecological Inference*. Beverly Hills, California: Sage Publisher.
- Lazarsfeld, P. F., & Rosenberg, M. (Eds.) (1955). *The language of social research*. New York: Free Press.
- Lincoln, J. R., & Zeitz, G. (1980). Organizational properties from aggregate data: Separating individual and structural effects. *American Sociological Review*, 45, 391-408.
- Lindquist, E. F. (1940). *Statistical analysis in educational research*. Boston: Houghton Mifflin.
- Mason, W. M., Wong, G. Y., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological methodology 1983-1984* (pp. 72-103). San Francisco: Jossey Bass.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School Matters*. Berkeley: The University of California Press.
- O'Brien, R. M. (1991). Correcting measures of relationship between aggregate-level variables. In P. Marsden (Ed.), *Sociological methodology* (pp. 125-165). Washington, D.C.: American Sociological Association.
- O'Brien, R. M. (1990). Estimating the reliability of aggregate-level variables based on individual-level characteristics. *Sociological Methods and Research*, 18, 473-504.
- O'Brien, R. M. (1984). Using generalizability theory to estimate the dependability of aggregate-level variables. *Quality and Quantity*, 18, 193-205.
- Raudenbush, S. W. (1988). Estimating change in dispersion. *Journal of Educational Statistics*, 13, 148-171.
- Raudenbush, S. W., & Bryk, A. S. (1989). Quantitative models for estimating teacher and school effectiveness. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 205-232). San Diego: Academic Press.
- Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, 15, 423-475.
- Raudenbush, S. W., & Bryk, A. S. (1987). Examining correlates of diversity. *Journal of Educational Statistics*, 12, 241-269.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Rogosa, D. (1980). Time and time again: Some analysis problems in longitudinal research. In C. E. Bidwell & D. M. Windham (Eds), *The analysis of educational productivity, volume II: Issues in macroanalysis* (pp. 153-201). Cambridge, Mass.: Ballinger.
- Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.
- Stipak, B. (1980). Analysis of policy issues concerning social integration. *Policy Sciences* 12, 41-60.
- Stipak, B., & Hensler, C. (1982). The workshop: Statistical inference in contextual analysis. *American Journal of Political Science*, 26, 151-175.
- Stockard, J. & Mayberry, M. (1992). *Effective Educational Environments*. Newbury Park CA: Corwin/Sage.
- Tannenbaum, A. S., & Bachman, J. G. (1964). Structural versus individual effects. *American*

- Journal of Sociology*, 69, 585-595.
- Wiley, D. E. (1970). Design and analysis of evaluation studies. In M. C. Wittrock, & D. E. Wiley (Eds.), *The evaluation of instruction: Issue and problems* (pp. 259-269). New York: Holt, Rinehart and Winston.
- Wittrock, M. C., & Wiley, D. E. (Eds.). (1970). *The evaluation of instruction: Issues and problems*. New York: Holt, Rinehart and Winston.
- Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80, 513-524.
- Zeller, R. A., & Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge: Cambridge University Press.