Methodological Issues in Studying the Effectiveness of Schooling: Recent Developments and Lingering Concerns

Kenneth A. Sirotnik
Leigh Burstein

CSE Report No. 215
1983

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

The research reported herein was supported in part by a grant to the Center for the Study of Evaluation, from the National Institute of Education, U.S. Department of Education. However, the opinions and findings expressed here do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

Table of Contents

In the event that the title of our paper has resurrected the mythical separation of methodology and substance, we wish to vigorously eschew this dichotomy and affirm the notion that among the most serious "methodological" issues confronting educational research is the adequacy of the conceptualization of the phenomena. Measurement and analysis problems are crucial, but secondary and contingent, of course, on the prior issue. This is especially true in light of how seductive the recent sophisticated developments in quantitative technology have become and yet how vulnerable methods such as LISREL are to poor concepts, poor data, or both.

We wish to offer a constructive critique of the research agenda proposed by the Centre for Educational Sociology (CES; Cuttance, 1983). We will, therefore, be unable to avoid at least a couple of conceptual concerns. These, in turn have profound analytical consequences, most of which can be subsumed under the technical heading of "specification bias." Thus, we have organized what follows into separate but interrelated discussions, one conceptualizing the terrain, the other traversing it.

CONCEPTUALIZING THE TERRAIN

You remember the old saying: "You can't see the forest for the trees." Perhaps the adage has even been applied to you, as it has to us, on those occasions when our preoccupation with details has caused us to lose sight of the larger picture. But it also works the other way around. There have been many times that we have failed to see the trees for the forest. In our attempts to grasp the larger picture we have lost sight of the important features without which the picture becomes sorely attenuated. Studies

of school effects and school effectiveness can be (and have been) victimized by both versons of this danger in the woods. Up until the last half dozen years or so, such studies tended to focus exclusively either on macro variables (e.g., resource allocation) with ostensibly policy-oriented implications or micro variables (e.g., time on task) with ostensibly instruction-oriented implications.

However, recent trends in macro- and micro-analysis (see, for example, Bidwell & Windham, 1982; and Dreeben & Thomas, 1980) suggest an emerging awareness that both kinds of orientations are necessary to achieve any practical understanding of educational productivity and schooling in general. Failure to simultaneously take into account such features as district accountability procedures, principal management styles, instructional beliefs of teachers, classroom pedagogical practices, individual student differences in ability and attitude, parent support structures, and extra-school learning opportunities--to name just a few variables--can seriously under-represent the complexity and interactivity of the schooling process, thereby precluding even the possibility of determining any cause-effect explanations.

Notwithstanding the awareness of these issues exhibited by Cuttance (1983) in his discussion of the various theoretical frameworks used for studying school effects and schooling effectiveness, we feel that the agenda, as proposed, represents a step backwards--in many respects, both the trees and the forest are difficult to discern. To be sure, these sentiments are by and large traceable to the post hoc nature of this study. Cuttance is in the unenviable position of trying to creatively analyze

data collected by others for what appear to be conceptually different reasons. To simply admonish the proposed study for data it does not have, therefore, would serve no constructive purpose.

Unfortunately, this admonition is unavoidable in face of the incongruity between the limitations of the information available and the ambitiousness of the research agenda and analytical capabilities as advertised. A number of concerns come to mind here, but we will detail only two general categories: (1) the idea of school effects and effectiveness and the separation of inputs, processes, and outcomes, and (2) the thickness and breadth of information necessary and available for defining the effects of schools and schooling.

Effects, Effectiveness, and Plausible Causality

Inherent in the attempt to define school effects or the effectiveness of schooling is a more fundamental problem for all effects research. Given the time frame selected, one researcher's outcome may be another's input. Moreover, time alone does not necessarily provide the conceptual meat for arguing the plausibility of an antecedent-consequent relationship.

Usually one specifies a slice of time of interest. Then, subject to conceptual argument, variables are designated toward the end of this time interval as outcomes or effects of an intervening schooling process. Typically, these outcomes are taken to be one form or another of achievement measures and the effect is then manifested in achievement score variation. Less often, outcomes in the affective domain are also assessed and school effects can be manifested in the variance of these measures as well.

However, it is by no means a foregone conclusion that these outcomes are consequences in a causal chain. It may well be, for example, that achievement and/or affect are confounded with familial and/or social environments that set up expectations for schools which fulfill, by and large, the prophecy. In this sense, then, the effects may be, in part, their own antecedents. We raise this issue not with the expectation of resolving it but simply as a reminder that the risk of overinterpreting such findings is not low. This would seem to be especially the case when "outcomes" are post-schooling sentiments and behaviors, when "antecedents" are largely demographic, and when little or no schooling process can be modeled.

Moreover, it is important to distinguish between a school's effects and the effects of schools. Assuming there exists variance on potentially relevant antecedent and process variables in a single school, the effects (and thereby the effectiveness) of that school can be studied. But if interests extend further than what goes on within schools, then the focus shifts to relative comparisons of schools with one another. School (or better, perhaps, schooling) effects are now manifested in the variance of outcome aggregates (typically the mean) at the school level. Accounting for between school variance on these outcomes, then, becomes the goal of schooling effect studies. When schools are sorted out on some criteria into particularly high and low outcome groupings and correlates are sought for this dichotomy, we have what usually are termed school (or better, perhaps, schooling) effectiveness studies.

But schooling effects can also be manifested in school-to-school dif-

ferences in within-school analyses. In other words, to the extent that within-school patterns are not readily pooled across schools, we have evidence for potentially different sources of schooling effects from those typically assessed by aggregating to the school level. We can, of course, extend this designation of between- and within-school effects to between- and within-class effects and studies of classroom/teacher effectiveness, between and within district effects and studies of district effectiveness, and all implied nestings of these designs as well. The various classes of studies that have occurred under the rubrics of school effects and effectiveness can basically be distinguished by the attention they pay to between- and/or within-school effects and the degree of comprehensivity to which they tap into the potential schooling terrain.

Necessary Versus Available Information

Any attempt to map out this terrain is plagued by inconsistencies when forcing certain information to fit certain cell designations. Nevertheless, when used with the heuristic intentions behind the schematics in Figures 1 and 2, such maps can serve to highlight conceptual, methodological and practical implications of different forays into the educational domain. Figure 1 is a modified version of that used by Sirotnik and Oakes (1981) to describe the contextual features of schooling. Here, we have added two additional facets, general categories of data types and levels at which data can be aggregated. Combined with the original facets of data sources and data domains, this schematic should begin to suggest the depth of information that is potentially relevant to explaining (and perhaps even understanding) the schooling phenomenon. Moreover, the aggregation facet is

not just an analytical addendum. The fact that the data collected at, or aggregated to, different levels may mean different things (Burstein, 1980a, 1980b; Sirotnik, 1980) requires recognition in any substantive framework.

Figure 2 complicates matters by acknowledging the necessary time factor and the fact that much of the information mapped out in Figure 1 is not static. Even in Figure 2, however, it is necessary to chop out some time segment. We have chosen to represent the usual K-12 elementary and secondary educational time frame and the potential for pre-school input and post-secondary outcome information. Different study purposes will, of course, dictate different points of entry and departure.

We have not attempted to fill in data examples for each cell in Figure 1. Clearly, some cells are naturally empty; for example, cognitive and attitudingal data can not be directly defined or collected on non-human entities. Thus, cells like those created by the intersection of cognitive and attitudinal columns in the instructional domain with the classroom data source row are undefined. This is not to say, however, that such data cannot be created at the classroom level by aggregating responses, e.g., the cognitive and attitudinal cells at the class level for students represent this kind of information. But the crucial issue, again, becomes what, in fact, is being measured by these aggregates.

Now let's return to the agenda proposed in the CES study. At one point in the proposal following a list of past studies, it is noted, by contrast, that "the design of the present programme is of greater scope conceptually." This is a rather conspicuous overstatement in light of the fact that we

can find no signs of student, teacher, administration, parent, or observational data collected during and on the process of schooling. Rather, what we are able to discern in the way of available data includes a small selection of structural data at the school level, some demographic data and incoming verbal reasoning ability for students, and, as "outcomes," several measures of students' retrospective[1] perceptions of their schooling experience, a few post-school attitudes and employment prospects, and achievement scores upon leaving secondary school.

Again, we have no wish to set up a straw-man critique. Yet we are led to expect in the proposed research agenda that from data such as these we can (a) better understand what makes some schools more effective than others; (b) determine the dimensionality of outcomes; and (c) test competing schooling models based upon the dynamic interplay of organizational and/or institutional constructs vis-a-vis teaching and learning. We applaud the ambitiousness of this agenda. But the most constructive advice we can offer is that the agenda be revised in ways that are commensurate with the modesty of the available information. Despite the confirmatory stance of the proposal, we believe that this study is at best exploratory and that techniques such as LISREL should be used primarily for suggesting hypotheses for further study. It is in this vein, then, that we wish to further discuss some more technical and analytical issues.

TRAVERSING THE TERRAIN

Having mapped out at least a sizeable chunk of the forest and the trees, it is necessary to assess the methodological complexities and analytical

problems depending upon how one wishes to traverse this terrain. Among the many issues that could emerge, two general classes of concerns are particularly crucial: (1) the complexities introduced by multilevel analysis and (2) the viability of covariance structure modeling. Again, we will discuss these issues both generally and in relation to the CES proposal per se.

Levels of Analysis

One sign of progress in school effects research over the last few years has been the maturing awareness of the multilevel character of the educational process. A few years ago it seemed trite but necessary to remind researchers that since no level of the educational system was uniquely responsible for the delivery of and response to schooling, confining substantive questions to a single level had inherent misspecification (e.g., Rogosa, 1978; Burstein, 1980a).

In their own way, each of the alternative models of schooling reviewed by Cuttance (1983) reflects a multilevel perspective of the organizational structure of schools and the implication of this structure for educational effects. But, with rare exceptions (e.g., Barr & Dreeban, 1983; Bidwell & Kasarda, 1980; Brown & Saks, 1980, 1981), empirical investigations of school effects still fail to conscientiously mirror their multilevel conceptions. Measurement and analytical complications inherent with multilevel data may be partly to blame for the dearth of proper empirical application. When technically sophisticated school effects researchers debate the implications of "statistical dependence among observations within groups" and of how to disentangle effects from various levels, it is not surprising that those

more interested in the substantive issues shy away from the potential analytical complexities.

Nevertheless, there are several methodological aspects of the levels questions that warrant consideration in school effects research of the type Cuttance has proposed. We have already alluded to one measurement issue, the possibility of a construct shift across levels, earlier in the paper. Elsewhere we have argued (Belandria & Burstein, in press; Burstein, Fischer, & Miller, 1980) that indicators of socio-economic background can measure different constructs at the school level than they do at the student level. In other studies, aggregated perceptions of the individuals from an organizational setting (e.g., students within classrooms or schools, teachers within schools) apparently have taken on a different meaning from their individual-level counterparts (e.g., Capell, 1979; Sirotnik, 198).

What makes the construct shift phenomenon an important concern in school effects research is that its analytical consequences are most severe when the guiding conceptualization of pertinent constructs and their operationalization are weak. The careful selection of a diverse array of indicators from different levels can minimize the hazards of construct shifts. Unfortunately, the reason the problem occurs in school effects research in the first place is that investigators often try to stretch inadequate indicators to measure an initially incomplete set of constructs.

Another measurement concern is the difficulty of adequately representing within-school distributions of outcomes by a single school-level indicator of central tendency. Analyses of between-group (class, school, etc.) means can hide important difference in the within-group distribution of pupil out-

comes and schooling processes. Schools offering different arrays of educational services can have the same mean outcome yet vary in other characteristics of the school's distribution (e.g., variance, proportion above or below a certain standard) or in the within-school relationships of student background to student outcomes (as reflected in regression coefficients). In their theory of production technologies and resource allocation within classrooms and schools, Brown and Saks (1975) attribute this variation to the consequences of differences in the "tastes" of different educators for maximizing different outputs (e.g., to minimize the number of low performers or ensure that every student masters basic skills rather than maximize mean performance.

The importance of considering a variety of statistical indicators of group performance in the presence of possibly divergent educational goals should be self-evident. A number of investigators (e.g., Barr & Dreeben, 1983; Brown & Saks, 1975, 1980; Burstein & Miller, 1981; Burstein, Miller, & Linn, 1981; Cooley & Lohnes, 1976; Klitgaard, 1975; Lohnes, 1972; Sirotnik, 1980; Spencer, 1983; Wiley, 1970) have either suggested or empirically examined a variety of possible group-level indicators that better reflect within-group distributions.[2] The evolving consensus appears to point to greater analytical attention to subgroups of students within schools (e.g., low ability or status vs. high ability or status) and to the identification of clusters of schools which serve similar populations of students and apparently share the same goals ("tastes" in the Brown and Saks sense). Separate or comparative analyses of student subgroups within and across clusters of schools would then be conducted and interpreted.

Cuttance is apparently sensitive to the levels of analysis issues in school effects research that might arise in the CES study. We have already alluded to consideration of multilevel matters in his discussion of the alternative models of schooling. Cuttance also proposes to "experiment with alternative measures which capture more information about the distribution of within school outcomes" (1983, p. 18). We enthusiastically support this exploration and see it as a prominent feature of the study. Any lingering concerns we have about endorsing a thorough multilevel analysis in the study stem from the previously cited limitations of the measures of schooling processes and organization, on the one hand, and the fear that the analytical machinery that may predominate in the study is not well-suited for handling multi-level investigations. We shall have more to say on the latter point in the next section.

Covariance Structure Modeling Techniques

Many of the analyses in the CES Study will apparently be carried out withing a covariance structure modeling framework, most likely via LISREL (Jöreskog & Sorbom, 1980). In theory, there are clear advantages of covariance structure modeling over the single-equation regression models typically employed in analyses of school effects. Most notable are the facility (a) to incorporate multiple indicators of key constructs and thus estimate structural (latent variable) relationships, (b) to specify and estimate reciprocal effects among key clusters of variables, (c) to estimate several model equations simultaneously (e.g., those "explaining" test performance as well as enrollment in postsecondary education), and (d) to compare the structural relations across groups (e.g., are the relations of student background char-

acteristics to educational performance the same for both comprehensive and selective schools?) or across waves of data collection. Moreover, the limitations from the multivariate normality assumptions required by LISREL estimation are now apparently resolvable through the use of techniques for structural modeling with both polytymous and continuous variable (using Muthen's LACCI program).

As with most sophisticated analytical advances, however, the major strengths of covariance structure modeling are also the source of its major weaknesses. One serious and endemic problem is that the adequacy of these methods is inherently dependent on the quality of the model specification--both in terms of the limits of theory (i.e., which constructs are pertinent) and of operationalization through the measurements one collects. Bad theory and bad data are no less bad just because the analysis is sophisticated and complicated. On the contrary, these are obvious dangers from the appearance of sophistication when theoretical and empirical foundations are shaky.

There are simply no ways around the inadequacies in the available measures of school processes and organization available in the CES Study. Thus, comprehensive covariance structure models of school effects using these data are misspecified from the outset and there are no viable ways to patch up this shortcoming.

A second concern is that although the primary reason many investigators turn to covariance structure modeling is its ability to estimate complex models with multiple latent constructs and measurements, the practical reality is that the estimation procedures are often overwhelmed by the

sheer size and complexity of the models. In the search for estimable models which "fit" the data, many investigators simplify their models to such a degree that conceptual and psychometric degradation is too great to justify a claim that the "theory" has received an adequate test.

The difficulties in handling the multilevel aspects of schooling data exemplify the complexity constraints on the viability of covariance structure modeling techniques. Most attempts to apply LISREL in school effects studies have chosen either to ignore the multilevel structure or to estimate relationships at each level separately. In perhaps the first full-scale LISREL application in an investigation of school effects, Munck (1979) estimated strictly a pooled within-school structural model. Separate between-student and between-school LISREL analyses were conducted as part of the Sustaining Effects Study (Wingard, 1980). From the perspective of current views about multilevel analyses, these attempts are either incomplete or do not consider the inherent linkages and conditioning relationships across levels.

Two covariance structure modeling strategies that are perhaps more attuned to the multilevel structure of educational data have recently received some attention. Building on Schmidt's (1969) work on the hierarchical decomposition of components of covariance Wisenbaker and Schmidt (1979) and Wisenbaker (1980) developed modeling techniques which simultaneously estimate structural relations at two levels (between-school and within-school). It is unclear, however, whether their technique is sufficiently flexible to accommodate relationships <u>across</u> levels as can occur through the non-random selection mechanisms typically guiding school assignment.

Schneider and Treiber (1982) employ a two-stage procedure in which homogeneous clusters of groups (e.g., classrooms, schools) are formed (either logically or empirically) and then the within-cluster structural relations are compared using LISREL. Their approach is intriguing in that it is potentially sensitive to between-school (class) differences in organizational structure and educational policies that might lead to differences in the structural relations characterizing schooling effects.

If either the Wisenbaker-Schmidt or the Schneider-Treiber approach (or some as yet unspecified alternative) prove to be viable across a range of school effects investigations, then lingering concerns about multilevel complexities in covariance structure modeling analyses would dissipate somewhat. But, even with data bases rich in process data at pertinent levels of analysis, attempts to estimate comprehensive covariance structure models of the type proposed in the CES study are likely to flounder from the sheer magnitude of the task. There are simply too many possible constructs, too many variables and too many pertinent levels. Nevertheless, covariance structure modeling can make a contribution to the Scottish Study once the trappings of comprehensiveness are set aside.

RECOMMENDATIONS

Repeating the advice we offered above, the primary "methodological" agenda item for Cuttance and the rest of the CES staff working on the Scottish school effects study is to realign their analytical plans in accordance with a conceptual plan befitting the scope and breadth of their available data.

In this regard, we envision that covariance structure methods could be used at various stages in the research. Their facility for modeling factor structure among indicators and constructs is a case in point. The factor structure underlying the outcome measures (both between and within schools, over time, over types of schools (e.g., comprehensive, selective) can be productively examined using covariance structure techniques, so long as the caveats we have sounded regarding these outcomes (see footnote 1) guide interpretations. Similarly, covariance structure techniques might be used to compare the within-school relationships of student background to outcome measures for different clusters of schools (defined a priori or empirically) within a given cohort or across cohorts. This latter type of study might follow the Schneider-Treiber (1982) strategy. In any event, the analyses must proceed in exploratory stages and not attempt to combine all possible complications into one cure-all model.

Clearly, there are important contributions this study could make to the literature on schooling. It is useful to keep in mind, for example, that Scotland is relatively ethnically homogeneous and its educational system is relatively uniform, especially in its curriculum and thus its assessment procedures. As has been demonstrated over the years in the IEA studies, there is value in a comparative perspective on the functioning of schools. A study of schooling effects in Scotland, therefore, offers a useful contrast to the U.S., the home base for most of the effectiveness investigations.

Continuing in a positive vein, we hope that the CES study will set an example for past, present and future studies by <u>not</u> claiming to have

uncovered a set of explanations and understandings about why and how some schools are more effective than others. Rather, they will report that they have explored the sources of variance in one highly selective set of variables (that they have chosen to designate as outcomes) on the basis of selected other variables. As a result, they will have suggestions for possible schooling targets that have promising implications for further and more in-depth study.

DATA DOMAINS

AGGREGATION LEVELS	Data Sources:	Personal (Individual) D/B	C	A	B	Instructional (Classroom) A/S	C	A	B	Institutional (School) A/S	C	A	B	Societal (Schooling) A/S	C	A	B
Individual	Students																
	Teachers																
	Administrators																
	Parents																
Class	Students																
	Teachers																
	Administrators																
	Parents																
	Classroom																
School	Students																
	Teachers																
	Administrators																
	Parents																
	Classrooms																
	School																
District	Students																
	Teachers																
	Administrators																
	Parents																
	Classrooms																
	Schools																
	District																

Data Categories:

D/B = Demographic &/or biographic
A/S = Archival &/or situational
C = Cognitive (typically CRTs &/or NRTs)
A = Affective (beliefs, attitudes, sentiments, etc.; typically collected via survey &/or interview questionnaire)
B = Behavioral (typically collected through observational methods)

FIGURE 1

A Partial Map of the
Educational Terrain

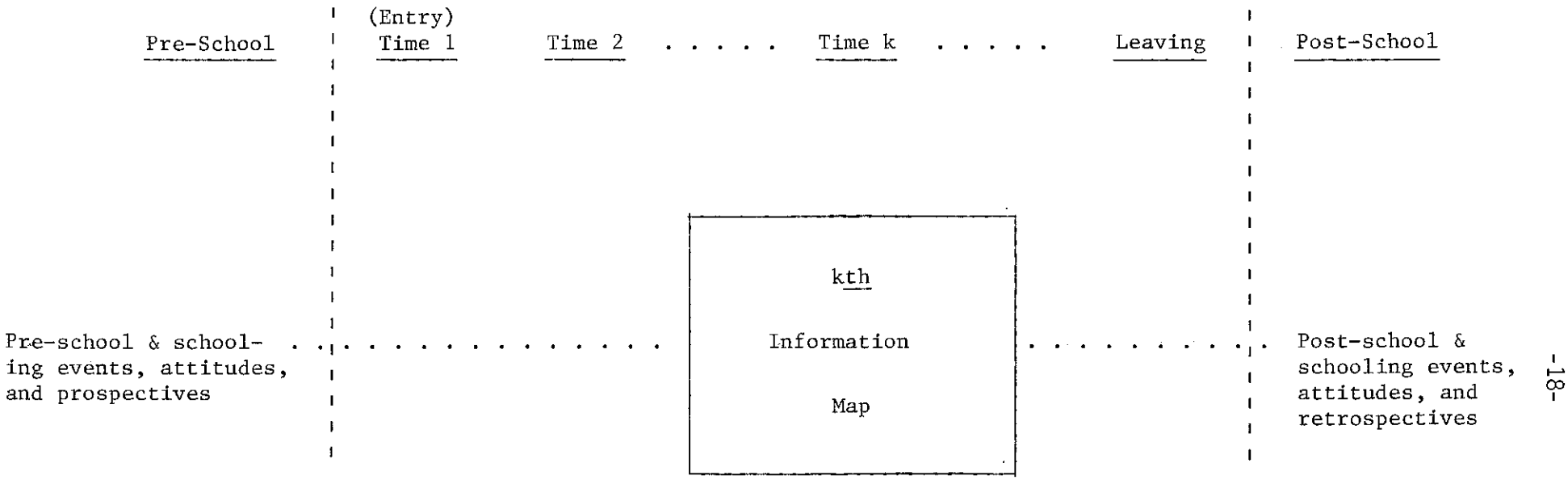

FIGURE 2

A Time Perspective

FOOTNOTES

[1] Cuttance is correct in noting that the CES study has more variety in designated "outcomes" than typically found in other studies. However, they are all tapping into the post-schooling domain. Moreover, a number of these outcomes in the affective domain require retrospective perceptions of attitudes/sentiments that can be gathered during schooling at the various times when such perceptions are relevant outcomes and inputs. The question (and a serious one) here is how these retrospectives are to be conceptually treated. If they are meant to be substitutes for in situ perceptions, then we are highly critical of their utility, especially from a modeling perspective. If, however, the retrospectives are treated in their own right as possibly important post-secondary outcomes, then we have no problem with their interpretation.

[2] Attempts to investigate within-school distributions of outcomes can be very tricky (see e.g., Spencer, 1983). The statistics likely to be employed in such analyses (standard deviations, regression coefficients) do not necessarily have the same distribution properties as means and thus are potentially difficult to use in complex covariance structure models.

References

Barr, R., & Dreeben, R. How schools work: A study of reading instruction. University of Chicago Press, 1983.

Belandria, L., & Burstein, L. Cross-national comparison of the multilevel effects of background on achievement: An extension of earlier findings. In H. Osthoek and P. van den Eeden (Eds.), Multilevel aspects in the educational process. London: Gordon and Breach, in press.

Bidwell, C. W., & Kasarda, J. D. Conceptualizing and measuring the effects of school and schooling. American Journal of Education, 1980, 88, 401-430.

Bidwell, C. E., & Windham, O. Analysis of educational productivity, Volume II: Issues in macroanalysis. Cambridge, Massachusetts: Ballinger Press, 1982.

Brown, B. W., & Saks, D. H. The production and distribution of cognitive skills within schools. Journal of Political Economy, 1975, 83, 571-593.

Brown, B. W., & Saks, D. H. Production technologies and resource allocations within classrooms and schools: Theory and measurement. In R. Dreeben & J. A. Thomas (Eds.), The analysis of educational productivity, Vol. I: Issues in microanalysis. Cambridge, MA: Ballinger Press, 1980.

Brown, B. W., & Saks, D. H. Microeconomics of schooling. In D. C. Berliner (Ed.), Review of Research in Education, Vol. 9. Washington, D. C.: American Educational Research Association, 1981.

Burstein, L. Analysis of multilevel data in educational research and evaluation. In D. Berliner (Ed.), Review of Research in Education, Vol, 8, 158-233. Itasca, Illinois: F. E. Peacock, 1980a.

Burstein, L. The role of levels of analysis in the specification of educational effects. In R. Dreeben & J. A. Thomas (Eds.), Analysis of educational productivity, Vol. 1: Issues in microanalysis, 119-190. Cambridge, MA: Ballinger Press, 1980b.

Burstein, L., Fischer, K., & Miller, M. D. The multilevel effects of background on science achievement at different levels of analysis: A cross-national comparison. Sociology of Education, 1980, 53(4), 215-255.

Burstein, L., & Miller, M. D. Regression-based analyses of multilevel data. In R. F. Boruch, P. M. Wortman, D. S. Cordray, and Associates (Eds.), Program Evaluations. San Francisco: Jossey-Bass Publishers, 1981, 194-210.

Burstein, L., Miller, M. D., & Linn, R. L. The use of within-group slopes as indices of group outcomes. CSE Report No. 171. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Capell, F. Interpreting multilevel data from program evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April, 1979.

Cooley, W. W., & Lohnes, P. R. Evaluation research in education: Theory principles and practices. New York: Irvington Publishers, Inc., 1976.

Cuttance, P. F. An agenda for research into school effects and the effectiveness of schooling in Scotland. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, 1983.

Dreeben, R., & Thomas, J. A. (Eds.), Analysis of educational productivity. Volume I: Issues in microanalysis. Cambridge, MA: Ballinger Press, 1980.

Jöreskog, K. G., & Sorbom, D. LISREL V: Analysis of linear structure relationships by maximum likelihood and least squares methods. Research Report 80-5. Department of Statistics, University of Uppsala, 1980.

Klitgaard, R. W. Going beyond the mean in educational evaluation. Public Policy, 1975, 23, 59-79.

Lohnes, P. Statistical descriptors of school classes. American Educational Research Journal, 1972, 9, 547-556.

Munck, I.M.E. Model building in comparative education. IEA Monograph Series, No. 10, Stockholm, Sweden: Almqvist and Wiksell International, 1979.

Rogosa, D. Politics, process, and pyramids. Journal of Educational Statistics, 1978, 3(1), 79-86.

Schmidt, W. C. Covariance structure analysis of the multivariate random effects model. Unpublished doctoral dissertation, University of Chicago, 1969.

Schneider, W. G., & Treiber, B. Classroom differences in the determination of achievement changes. Unpublished manuscript, 1982.

Sirotnik, K. A. Psychometric implications of the unit-of-analysis "problem" (with examples from the measurement of organizational climate). Journal of Educational Measurement, 1980, 17, 245-281.

Sirotnik, K. A., & Oakes, J. A contextual appraisal system for schools: Medicine or madness? Educational Leadership, 1981, 39, 164-173.

Spencer, B. D. On interpreting test scores as social indicators: Statistical considerations. Journal of Educational Measurement, 1983, 20(4), 317-333.

Wiley, D. E. Design and analysis of evaluation studies. In M. C. Wittrock & D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart & Winston, 1970.

Wingard, J. Analysis plan for technical report 15 from the study of the sustaining effects of compensatory education of basic skills. Santa Monica, CA: System Development Corporation, October, 1980.

Wisenbaker, J. M. Structural equation models applied to hierarchical data. Unpublished doctoral dissertation, Michigan State University, 1980.

Wisenbaker, J., & Schmidt, W. The structural analysis of hierarchical data. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April, 1979.