DOCUMENT RESUME

ED 108 984 SO 008 327

AUTHOR Burstein, Leigh
TITLE Data Aggregation in Educational Research: Applications.
PUB DATE 75
NOTE 60p.; A paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., April 1975)

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS *Data Analysis; Data Collection; Data Processing; *Educational Research; *Evaluation Methods; Predictive Measurement; *Research Methodology; Research Problems; *Statistical Analysis; Statistical Data

ABSTRACT

Since problems associated with the statistical methodology of educational research are becoming increasingly important, this paper examines a subset of problems associated with the analysis and interpretation of aggregated data. Two major questions arise: (1) if a researcher knows the level (e.g., individual, teacher/classroom, school, school district) at which inferences are desired, what complications arise from analyzing data at different levels? and (2) are there general guidelines for determining the appropriate units of analysis in a given research context? Five research contexts in which group observations can be used to estimate relationships among measurement of individuals are examined including contexts with missing observations, fallibly measured variables, the economy of analysis, anonymously collected information, and ecological inference. In choosing units of analysis, appropriateness is a function of the questions asked and of the sampling and/or experimental unit. The former is reflected in the conceptualization of the research objective while the latter can indicate the presence of statistical constraints on the level of inference. Examples of issues and problems that arise with each concern are provided. (Author/DE)

SESS # 27:15

ED108984

SE 008 327

DATA AGGREGATION IN EDUCATIONAL RESEARCH: APPLICATIONS*

Leigh Burstein[1]
University of Wisconsin-Milwaukee

In the last ten to fifteen years, the character and focus of educational research has changed. A key indicator of the present trend is the greater frequency of large-scale educational surveys and investigations (Project Talent, Coleman Survey, National Assessment of Educational Progress). A primary impetus for this transformation is the expanding propensity of the representatives of the various social science disciplines (particularly sociology and economics) to attempt to detail the complex phenomenon called "schooling". The effects of schooling are what Coleman and his colleagues (1966) sought to elucidate, and what their critics and supporters (e.g., Mood, et al (1970), and Mosteller and Moynihan (1972)) attempted to clarify, deny, or deify. The topic was worth three volumes (plus massive technical appendices) by Mayeske and his colleagues (1972), eleven volumes by International Association for the Evaluation of Educational Achievement (IEA) (1967, 1973, 1975), and a catalogue of studies and summaries by Averch et al (1972).[2] The attempts to evaluate the major educational innovations of the War on Poverty also adopted the mode of analyzing school effects data that are massive both in terms of persons and characteristics measured.

Unfortunately, one does not expand from investigations of a few classrooms to nation-wide or even cross-national studies of schooling without encountering new and perhaps novel complications. And, if it is difficult to adequately control even a single-class, short-term experiment (and the evidence indicates that it is), how can we hope to maintain

*Paper presented at the Annual Meeting of the American Educational Research Association, April 3, 1975, Washington, D.C.

controls for 500 classrooms or 200 schools or 10 countries? The answer is that we simply cannot, but it is believed that the broader perspective afforded by the more "macrocosmic camera" yields a more realistic, better generalizable image of the phenomenon of schooling.

Moreover, large-scale investigations of schooling enter the realm of socio-politics and generally become the instrument for policy analysis. This, too, is a double-edge sword -- the deathknell of abstract theory and the awakening of socio-politico-economic consequences. Thus today we find ourselves as much in need of developing our political skills as we do our research capabilities. Given present political contingencies, there is a greater burden on the researcher to ensure that his analytical procedures and data interpretation are conducted in a manner that can withstand substantive professional criticism.

Despite the introductory remarks, this presentation is not about the philosophy or politics of education per se. My primary concern is with a subset of the problems associated with an increasingly important aspect of the methodology of educational research -- the analysis and interpretation of aggregated data. Aggregated data are encountered in almost all large-scale educational studies simply because schools are aggregates of their teachers and pupils, and classrooms are aggregates of the processes and persons within. The grouping of data can be simply a modest attempt to pare research costs and/or "scrub" dirty data, and in these instances, aggregation has relatively innocuous consequences. The use of aggregated data can enhance or obfuscate efforts to identify the relations among measures of human behavior. Often, the social and political context of the investigation will determine whether data aggregation occurs and whether interpretations based on aggregated data are enlightening or illusionary.

In an earlier document (Burstein, 1974b), three layers of problems related to aggregating data in educational research were identified:

(a) problems in grouping of observations or change in units of analysis;

(b) problems in cross-level inference, which is better known as the identification and analysis of contextual effects;

and

(c) problems in determining the appropriate units of analysis.

Here we shall focus on the evidence from research on "change in units of analysis" problems and the overriding issue of appropriate units of analysis. The issues related to cross-level inferences have been considered in the Division G roundtable on "Contextual Effects" and we will not attempt to elaborate on that discussion.

DEFINING THE ISSUES

Problems of data aggregation have important implications for educational researchers who are interested in relations among observations on individuals. For instance, the investigator may want to know the coefficient from the regression of student achievement on other student characteristics. However, these measurements cannot always be examined at the individual level. The data may not be obtainable or identifiable for each person, because of intact school or classroom reporting, or the school or classroom may be the sampling unit, or it may be too costly to analyze data at the individual level. Faced with such problems, observations on individuals are grouped according to, say, classrooms (schools) and between-group (e.g., classroom, school) regression coefficients are calculated. The investigator then may attempt to make inferences about the relations among individuals from the results of the analyses at the group level.

Given a situation like the one described above, two major questions arise:

(1) If we know the level (e.g., individual, teacher/classroom, school, school district) at which inferences are desired, what complications arise from analyzing data at different levels? ("Change in the units of analysis" or change-in-units problems.)

(2) Are there general guidelines for determining the appropriate units of analysis in a given research context? ("Appropriate units of analysis" or appropriate-units problems.)

Much is already known about change-in-units problems (Burstein, 1974a, 1975; Hannan and Burstein, 1974). The latter question subsumes changes-in-units problems. In general, however, the issues surrounding appropriate units in the social sciences are presently conceptual and are dealt with substantively on a case-by-case basis. We offer guidelines below that will hopefully clear up the blatant errors in selecting appropriate units and suggest ways of proceeding when the appropriate choice of units is not obvious.

CHANGE IN THE UNITS OF ANALYSIS -- THE GROUPING OF OBSERVATIONS

In general, complications can arise in translating relations from one level of analysis to another. Our primary concern is with change-in-units problems where the relations at the level of individuals are of interest, but the data are aggregated over individuals.

The degree of investigator control over the aggregation of data is a primary determinant of the complications due to grouping. In certain contexts, group membership is determined in some natural way, e.g., school attended, or census tract, and is thus beyond the investigator's control except for exclusion of sampling units and individuals (limited or no investigator control). In other contexts, the investigator can manipulate the formation of groups, either completely or partially. There are generally more options in the latter contexts for improving estimation.

Research Contexts

We can identify five research contexts in which group observations are used in estimating relations among measurements on individuals. These contexts include problems with (A) missing observations; (B) fallibly measured variables; (C) the economy of analysis; (D) anonymously collected information; and (E) ecological inference. The degree of investigator control over the formation of groups varies according to context. There are also differences among contexts in the reasons why the methods of data aggregation are used, how such methods are applied, and where they are principally applied. Table 1, reproduced from Burstein (1974a), summarizes the characteristics of each context.

Insert Table 1 here

Complete Investigator Control. In the first three contexts, the investigator has considerable flexibility about the choice of grouping methods. However, the problems addressed in context (A) (Kline et al, (1971) and (B) (Blalock et al, (1970) have seldom been subjected to aggregation procedures as other statistical methods are considered more suitable. (See Affifi and Elashoff (1966, 1967) on the missing observations problem: and Madansky (1959), Blalock, et al (1970), Blalock (1971), and Wiley and Wiley (1971) on the measurement error problem.) The procedures for selecting a suitable aggregation procedure are the same as in Context (C) and will be discussed along with that context.

Sound principles have already developed and demonstrated for data aggregation where the size and economy of analysis (Context C) is the

Table 1. Research contexts for data aggregation.

Research Context	Reasons for Use of Data Aggregation	Description of Application	Principal Application
I. Complete Investigator Control -- Group membership can be defined by any characteristic in the data set which is measured for all individual-level units.			
(A) MISSING OBSERVATIONS	Missing observations on primary variables for some individuals inhibit confidence in analytical results.	Each missing observation on a primary variable is replaced by the mean response on that variable for some group to which the individual belongs.	Longitudinal and cross-sectional analysis of survey data.
(B) FALLIBLY MEASURED VARIABLES	Random errors of measurement associated with independent variables attenuate estimates of regression coefficients.	Different approaches have been suggested as part of the general refinement of statistical procedures for handling "errors-in-variables" problems.	Statistical treatment of measurement errors in many studies.
(C) ECONOMY OF ANALYSIS	Budgetary constraints make analysis of massive data bases at the individual level impractical.	Data are collapsed into smaller administrative units by some grouping rule.	Analysis of census data and national, regional, and state school statistics.
II. Partial Investigator Control -- Group membership can be defined by any characteristic which has been measured simultaneously with each primary variable.			
(D) ANONYMOUSLY COLLECTED INFORMATION	Data on certain primary variables is collected anonymously making it impossible to match observations on primary variables at the individual level.	Characteristics measured simultaneously with both anonymously collected and identifiable information can be used to aggregate the data.	Confidential student records and responses to attitudinal questionnaires.

Table 1 (continued). Research contexts for data aggregation.

Research Context	Reasons for Use of Data Aggregation	Description of Application	Principal Application
III. <u>Limited or No Investigator Control</u> -- Group membership is determined prior to the collection and analysis of data; group membership is directly pertinent to the study of primary variables.			
(E) ECOLOGICAL INFERENCE	The sampling units of the investigation constitute "natural" aggregates of individuals.	Disaggregation efforts are generally a necessary precondition to reasonable inferences at the individual level.	Analysis of school and classroom means where the school and the class are the sampling units; data organized by census tract or demographic region.

concern (Prais & Aitchinson, (1954): Cramer, (1964); Feige and Watts, (1972): Hannan and Burstein, (1974); Burstein, (1974a). Compression of data is an issue when a large amount of data is collected and it can be reasonably concluded that a substantial savings in research costs can be obtained with minimal information loss. As data sets such as the 1,000,000 student Coleman Survey (Coleman, et al, (1966)) and the 300,000 case IEA Six-Subject Survey (e.g., Comber & Keeves, (1973)) becoming increasingly common in educational research, the investigator must clearly weigh the merits of data aggregation as opposed to initial restrictions on sampling.

Sampling limits data collection and cuts back on costs from the start, but leaves little possibility for extension beyond what is collected. Aggregation approaches require large collections of data which can be more costly if the information did not have to be collected in the first place.[3] However, the reduction in analysis costs through data aggregation can be substantial, and those aspects of the research that require it can still be conducted at the individual level. Thus, a "correct" choice is not always available and one must decide what considerations are most important.

Data aggregation procedures can also be employed in the analysis and reporting of confidential information. When researchers analyze Census data reported by classifications such as "years of education" and "ethnicity", they are, in effect, examining relations involving aggregate measures. In studies like those with Census data,[4] information on individuals is available and perhaps personally identifiable. However, practice and/or statutory considerations dictate that reporting be done on some aggregate basis. Thus, confidentiality of information on individuals is protected by the reporting process.

The procedures for identifying accurate estimates of individual parameters from data aggregated to maintain confidentiality are the same as in Context (C). As long as individual-level information is collected in an identifiable manner, the investigator can use his knowledge of the individual-level relations of primary variables to the method of grouping to identify a procedure that will be particularly suitable for minimizing information loss through grouping. Our examples in Appendix A demonstrate how guidelines based on the "structural equations" approach (Hannan (1971); Hannan and Burstein (1974); Burstein (1974a)) and statistics developed by Feige and Watts (1972) can be used to identify grouping methods with minimal information loss in this context.

Anonymously Collected Information. The use of aggregation techniques for analyzing anonymously collected information (Context (D)) is a relatively new notion.[5] What distinguishes this context from the others is that it is impossible to match observations on all primary variables at the individual level because information on certain primary variables has been collected anonymously. An application for grouping in this context would be in a study of the relations between student achievement and student attitudes where attitude data has been collected anonymously.

In order to use grouping methods in Context (D), the investigator collects information on potentially suitable grouping characteristics in addition to variables of primary interest. The individual observations are then collapsed into different groups and the parameters of interest can be estimated from the between-group relations. This procedure is viable as long as the potential grouping characteristics are measured simultaneously with each primary variable regardless of whether the information on the primary variable has been collected anonymously or with the individual identified. The grouping characteristics must also satisfy certain conditions necessary for precise estimation for all change-in-units problems. In Appendix B we present examples where individual-level relations are estimated from aggregation procedures that can be employed when some data have been collected anonymously.

We recognize that conducting research on confidential data presents very complicated social and political problems. (See Boruch, (1971a, 1971b, 1972a, 1972b)). There is a definite need for the privacy and protection of subjects in social research, but in many cases, limits on individual confidentiality are necessary for arriving at a better understanding of the socio-cultural milieu. Educational research is no exception to this dilemma. There are many so-called facts based on confidential data which would disappear with the more adequate qualifiers found in individual-level data. Robinson's (1950) paper provides an excellent example of an aggregation-induced misinterpretation.

The procedures suggested in this paper offer individuals assurances of their anonymity, while maintaining the possibility of carrying out research on topics that can further understanding of the complex interactions among individuals and institutions. Our basic premise in suggesting aggregation procedures in this context is that individuals can be protected through analysis methods which allow examination of relations among human characteristics without directly identifying the participating individuals.

<u>Ecological Inference</u>. The topic of ecological inference (Context (E)) has been extensively discussed in the sociological literature (See Robinson, (1950); Menzel, (1950); Duncan and Davis, (1953); Goodman, (1953, 1959); Scheuch, (1966) includes a particularly cogent discussion of the problem). Earlier debates centered around methods for overcoming the "ecological fallacy" of using areal data (e.g., data aggregated by census tract or

state) to estimate relations among characteristics of individuals. The dramatic change cited by Robinson (1950) in the size of the correlation between illiteracy and race as a function of the units of analysis (.95 at the regional level, .77 at the state level, .20 at the individual level) warn us not to glibly assume that group data provide the same information as data on individuals.

Educational researchers have seldom considered the potential "ecological" fallacy of inferring relations between properties of individuals from ecological (between-group) correlations. Any examination of recent research on the effects of schooling will indicate that many investigators perform between-group analyses, often without even addressing the question of whether the relations estimated at the group level are applicable at the individual level. Properties of students have been aggregated to the classroom level (e.g., Walberg, (1969)), the school level (e.g., Burkhead, Fox and Holland (1967); Hanushek (1968); Katzman (1963)), the school district level (Kiesling, (1969, 1970); Bidwell and Kasarda (1975)), the college level (e.g., Rock, et al. (1970); Baird and Feister (1972)), and even the state level (Walberg and Rasher (1974))!

Teacher characteristics are often aggregated to the school level even when information on individual students is used in the analysis. This was the case in the Coleman Report (Coleman, et al. (1966)) and all of its re-analyses (See, e.g., Levin (1970), Michelson (1970), and Smith (1972)), and it was also true with the IEA studies (Husen et al, (1967); Combe and Keeves (1973); Purves (1973); Thorndike (1973); and others). Thus it is impossible in these studies to match individual students with data on their own teachers. Under these circumstances, it is unlikely that school resources measures can account for a large proportion of the variation in individual student achievement since the variation in achievement among students within the same school accounts for the most of between-student variation.

In the face of the potential complications from data aggregation, the track record for input-output studies of the effects of schooling is not very good. In a report on school effectiveness research for the President's Commission on School Finance, Averch et al (1972) reviewed the major input-output studies to date. Of 19 studies cited, the school was the unit of analysis 6 times; the school district, 5 times; and in 7, school and teacher variables entered the analysis as school-level aggregates, while student variables were entered at the individual level. Only a study conducted by Hanushek (1970) matched individual students with their own teacher. Hanushek's (1972) followup on his earlier work and a recently reported investigation by researchers at the Federal Reserve Bank of Philadelphia (Summers and Wolfe (1974)) are the only other major studies with students and teachers matched.

All three studies using matched student-teacher data were based on a single school district and thus are not necessarily comparable to the larger studies that use schools or school districts as the units of analysis. We need more such studies involving multiple school districts before we can adequately grasp how much misinformation has been generated by the between-school and between-school district analyses.

The point is that if we agree with Averch et al (1972) that "the researcher would like to examine the relationship among school resources an individual student receives, his background, and the influences of his peers on one hand and his educational outcome on the other (page 38)", then we need to use individual-level data or at least be assured that the aggregation process does not distort the relations among important variables. At this point it appears to be impossible to avoid distortion through aggregation in this context. In fact, at present, we can not be sure that the variables that appear at one level of analysis will necessarily appear at other levels.

Data from the IEA studies of the factors influencing educational achievement can be used to illustrate the extent of our present state of ignorance.[6] This study involved twenty-one countries, considering six subject areas (Science, Reading Comprehension, Literature, Civics Education, English as a Foreign Language, and French as a Foreign Language) at three age levels (basically, 10 year-olds, 13 year-olds, and students in their pre-University year). Over 700 student, teacher, and school characteristics were measured.

Among the myriads of analyses and sub-investigations in the IEA studies were between-student and between-school regression analyses with achievement measures as outcome variables. The independent variables included home background factors (plus sex and age) (Block 1), type of school and program indices (Block 2), school and teacher variables (Block 3), and other student measures (Block 4), entered in an ordered fashion. We focus here on the data from the United States for all age levels in Science, Reading Comprehension and Literature.

As stated earlier, the teacher measures were included as school-level aggregates in both levels of analyses. Thus we might expect that their contribution to the variance accounted for would increase relative to home background factors as the analysis shifted from the student level to the school level, simply because they are then on more "equal" ground with the other independent variables. But that is not the case as Table 2 clearly illustrates.

Insert Table 2

For all age levels in each subject area, the contribution of home background

Table 2. Percentage of variance accounted for by home background measures (Block 1) and recent learning conditions (Block 3) in the Between-Student and Between-School analyses in three subject areas (Science, Reading Comprehension, and Literature), at three age levels (10 year olds, 13 year olds, and pre-University year (17)), from the IEA study of educational achievement in the United States.

		Percentage of Variance Accounted for			
		Home Background Factors (1)[c]		Recent Learning Conditions (3)[d]	
Subject Area	Age Level	Between Student	Between School	Between Student	Between School
Science	10	176	668	088	079
	13	217	672	048	105
	17[e]	184	435	081	131
Reading Comprehension	10	198	629	039	062
	13	221	668	032	055
	17[e]	175	526	029	000
Literature[f]	13	185	549	077	082
	17[e]	163	346	050	223
Median		185	589	049	081

[a]These data are taken from a preliminary version of An Empirical Study of Education in Twenty One Countries: A Technical Report, by Gilbert F. Peaker, which is due to be published later this year. They are reproduced here with the permission of the International Association for the Evaluation of Educational Achievement (IEA) who sponsored the report. These figures should not be cited without checking with the original source.

[b]Decimal points have been deleted for table clarity. For example, 176 should be read as "17.6% of the variance is accounted for".

[c]The composition of Block 1, which was labeled "Home Background Factors", varies according to the particular subject area and age level but generally includes an composite index of father's occupation, mother's and father's education, and the presence of a dictionary in the home plus student sex and student age.

[d]The composition of Block 3, labeled "Recent Learning Conditions", varies according to the particular subject area and age level. The percentages of variance accounted for by Block 3 in the table above reflects their contribution after home background factors (Block 1) and indices of the type of school and type of program in which the student is enrolled (Block 2) have already been entered.

[e]The actual description of the third population of students is "all students in their pre-University year" presumably twelfth grade in the United States.

[f]Ten year olds were not included in the Literature study.

factors increases dramatically accounting for more than half of the between-school variation in achievement (compared to roughly 20% of the between-student variation) while the contribution of recent learning conditions increases only slightly in most cases.

The fact is that students tend to group together in schools with other students from similar home backgrounds (Comber and Keeves, 1973). Therefore, aggregating data to the school level does not greatly reduce home background variation. As a result, home background factors become more influential, rather than less so, when the unit of analysis shifts from the student to the school. It also means that it becomes harder for school resources variables to significantly contribute to variation in achievement as the home background variables have already accounted for a "larger chunk of the smaller pie" through their earlier entry.

The picture is further complicated by the fact the the recent learning conditions which appear to affect achievement variation in the IEA study are not necessarily the same in the between-student and between-school analyses. For example, the pupil/teacher ratio, the opportunity to observe experiments and the presence of a science teacher for 10 year-olds in Science, and the number of hours of homework per week reported by the student for 13 year-olds in the Science study make significant contributions in the between-student analysis but fail to appear in the between-school analyses. Examples of important variables identified in the between-school analyses that do not contribute in the between-student analyses are the use of audiovisual methods (10 year-olds in Science) and a composite measure of the student's perception of the school environment (13 year-olds in Science). And on and on and on . . ., the same lack of consistency in every subject in virtually every country.

If the IEA studies accurately reflect present knowledge, we obviously have a lot left to learn about aggregation over students and teachers within schools and within larger administrative units. The guidelines we prescribe later on and the examples we present suggest that it is possible under certain conditions to find reasonable estimates of individual-level relations from group data or to determine when grouped data yield highly biased estimates of individual-level parameters. So far however, the utility of these guidelines have been demonstrated only for the case where the characteristic that determines group membership (we call it the "grouping variable") is measured on at least an ordered scale (see Hannan and Burstein, (1974), Burstein (1974a)), which is not the case here.

We have some notions about how to proceed when nominal "ecological" variables such as school and school districts are used to form groups, but have been unable to obtain the necessary data to try out our ideas on any reasonable scale. The ideal data set for such an investigation contains all variables measurable at their lowest possible level -- individual students matched with their own teachers and characteristics of their own school setting for students from multiple schools and, hopefully, multiple school districts. Then we would do the following:

1. Individual Model
   a. Determine the best model for school effects at the individual level;
   b. Aggregate inputs from the model (independent variables) successively to the level of classroom, school, and school district;
   c. Aggregate both outcomes and inputs to the classroom, school, and school district levels; and,
   d. Attempt to identify the factors that coincide with changes in coefficients across levels.

2. Classroom Level

   a. Find the best model for school effects at the classroom level;

   b. If different from the model at the individual level, aggregate and/or disaggregate inputs to the level of the school, school district, and the student;

   c. Aggregate and/or disaggregate both outcomes and inputs to the level of the school, the school district, and the student.

3. Repeat #1 and #2 starting with the best model at the school level.

4. Repeat #1, #2, and #3 starting with the best model at the school district level.

What we would expect to find is that (i) there are substantial differences in the magnitudes of the coefficients across levels for a given model, (ii) different variables enter the models at the different levels, and (iii) the coefficients for variables that appear in multiple "best" models will differ across models even at the same level of aggregation. The only conditions that would lead us to expect results other than the above would be if schools (classrooms, school districts) were random groupings of students and teachers, or if there were no between-school (classroom, school district) variation in performance on outcome measures that could not be accounted for by input variables in the "best" models. Neither of these conditions is likely to occur.

The ideal study described above has not been done and probably cannot be done with any presently available sets of educational data. The closest thing to it so far is the Hannan, Freeman, and Meyer analysis (1975) using the Mercer data which lacks the match between students and their teachers. They found the expected changes in the models across levels and large differences in the coefficients for the various input variables (Hannan discusses their work further in his own presentation (Hannan (1975)).

CHOOSING THE APPROPRIATE UNIT OF ANALYSES

In choosing a unit of analysis, the researcher is in effect making some choice with respect to the level of aggregation which suits his purposes. Whereas change-in-units are limited to selecting among alternative possibilities for compressing data or interpreting the information loss where the choice of grouping method is beyond the investigator's control, choosing the units of analysis involves the specification of a set research foci in addition to determining an aggregation-disaggregation scheme. In general, appropriateness is a function of (a) the question asked and (b) the sampling and/or experimental unit. The former is reflected in the conceptualization of the research objective while the latter can indicate the presence of statistical constraints on the level of inference. Below we discuss and provide examples of the kinds of issues that arise with each concern.

Conceptual Unit of Interest

The issue here is the question of what is the research objective. If observations are generated from an appropriate sampling design, the data at every level in the ideal study -- pupil, classroom (teacher), school, school district and so on -- can be used to investigate certain empirically based questions. Though aggregate data are usually inappropriate for studying properties of individual members of the aggregates, the investigator may be interested in the behavior of the aggregate units themselves.

<u>Classrooms as Units</u>. Studies of group process in educational settings can be meaningfully conducted at the level of the classroom. For example, though the methodology is not entirely satisfactory, Walberg (1969) appropriately concentrates on the classrooms as the unit of analysis in studying classroom climate. Walberg's purpose is to "replicate the work on the effects of classroom climate on learning and to investigate . . . effects of student biographical characteristics . . . on learning for the <u>class as a whole</u> (Walberg (1969), p. 529, emphasis added)". Apparently, he has no interest in applying his findings to individuals.

<u>Schools as Units</u>. There are many situations in which the school can be the appropriate unit of analysis, though most of the studies cited in Averch <u>et al</u> (1972) do so incorrectly. If the purpose of the investigation is to identify unusually "effective or ineffective" schools or to depict differences in practices among schools, between-school analyses are called for. A Rand sponsored study by Klitgaard and Hall (1973) contains a particularly thorough treatment of the process of identifying effective schools.

We (Burstein, Kremer, and Gemoll, in progress) are currently analyzing 3 years of school-level data which includes achievement, student background, school resources, and teacher experience indices. Our purpose is to identify those schools which appear to be unusually effective or ineffective in terms of proportion of low scoring and high scoring students after controlling for student background characteristics (ethnicity, mobility, etc.). Once such schools are identified, the school resource and teacher training data will be examined and further information will be gathered in a case-study fashion on each effective (ineffective) school.

The exercise described above might be conducted by the research and planning staff of a system in order to determine if there are cost effective ways to allocate the limited discretionary portion of school funds. This type of study is also cost effective in that the research staff expends most of its energy intensively investigating that subset of the school system that offers the highest potential payoff.

An unusually enlightening (from the point of view of appropriate units) series of studies has been conducted by research psychologists at ETS on various aspects of college environments. The college was the primary unit of analysis in a study by Rock _et al_ (1970) of the relationship between college characteristics and student achievement; by Rock _et al_ (1972) of the interaction of college effects and student aptitude (we return to this article later on), and by Baird and Feister (1972) of grading standards.

The Baird-Feister study exemplifies the perspective of the whole series. The overriding question in Baird and Feister's investigation was whether college grading standards are affected by changes in the abilities of entering students. As part of the work they sought to determine whether the same grade reflected the same level of student performance from college to college and from year to year in the same college. They consistently and appropriately examined college-level behavior in order to answer their questions.

School Districts as Units. For the most part, the conditions under which one might contemplate treating the school district as the unit of analysis parallel those of the school as unit. The research division of a state department of education might try to identify the characteristics of unusually effective or ineffective school districts in the same fashion as we are doing with schools. There is again a potentially substantial savings in research costs and perhaps the results can provide guidelines for cost-beneficial targeting of state funds.

Another instance in which the school district is the desirable unit is in studies of the administrative intensity and the like. A recent study by Hannan and Freeman (1975) demonstrates this usage of between-district data.

We could carry this line of inquiry on to higher levels, such as states and countries, and always be able to identify some question where the appropriate unit is a specific collective property. The mapping of appropriate units to answer specific questions is invariably logical. It is only in using a unit that is illogical at face value where problems generally arise.

For example, though their conclusions are appealing, it is the height of folly to believe that Walberg and Rasher (1974) have avoided the methodological shortcomings of the Coleman study by using state-level data, especially with 1969 and 1970 selective service examination failures as the outcome variables! Walberg had previously cited (if not read) Robinson's (1950) paper (Walberg, 1969, p. 530) and should know better than to treat states as random groupings of persons. Besides, anyone who was eligible for the draft in 1969-70 can tell you that (a) women did not take military service mental tests and (b) the years 1969-70 were not the times to try to pass the test (a case of the winners are the losers, in this writer's opinion).

Statistical Considerations

The important statistical considerations are of two types -- those related to sampling and experimentation and those related to model specification. These problems receive attention in two entirely different literatures, the former being of concern primarily to statisticians and psychologists, and the latter to econometricians and sociological "modelers".

Sampling and the Experimental Unit. The problems with sampling and the experimental unit represent familiar terrain. In experiments where sampling units are groups of individuals (e.g., classrooms), between-group analyses must logically be conducted even when the relations among measurements on individuals is of primary concern. The investigator lacks control over group membership and is therefore unable to determine how the required grouping procedure affects variation and covariation of the study variables. Under these conditions, the possibility of inferring relations at the individual level is limited.

In any case, the sampling of groups can present a particularly complex type of aggregation problem, since questions regarding sampling bias arise in addition to concerns about level of inference. For instance, the investigator needs to know whether the sampled classrooms are representative of classrooms in the universe to which he wishes to generalize in order for the between-group analyses to make sense.

Another facet of the sampling problem is the question of experimental unit. The experimental unit is "the smallest . . . collection of experimental subjects that have been randomly assigned to different conditions . . . and have responded independently of each other for the duration of the experiment (Glass and Stanley (1972), p. 506, emphasis added)". Thus, according to Glass and Stanley, studies involving intact classrooms should legitimately analyze classroom means. (Lewy (1972) states basically the same point.)

Glass and Stanley's arguments are sobering. In the extreme, they mean that between-student analyses are never appropriate with data provided from classrooms even if students were randomly assigned and sampled.

Our only salvation would be the nuances of random grouping and its effect on the relation of between-group to between-student analyses (see earlier discussion). Since most studies of the effects of schooling are at best quasi-experimental, the question of experimental unit is just one more methodological hurdle.

Specification Bias. We will not go into detail here as there is already substantial literature on the topic of specification bias and its interrelation to aggregation bias (Hannan (1971); Feige and Watts (1972); Hannan and Burstein (1974); Hanushek, et al (1974); Burstein (1975)).

The argument goes like this. The inability of social scientists to base their analyses of statistical models of human behavior on a well-defined theory often leads to misspecification through the inclusion of redundant (collinear) regressors or, more importantly, through the exclusion of causally relevant measures. Model misspecification as described above affects the relationship of a sample estimate of a coefficient to its population value.

Aggregation bias is a form of specification bias in that it arises through a lack of independence between the variables in the model and their disturbance terms. Furthermore, data aggregation can only exacerbate the problems caused by other forms of specification bias (Hannan and Burstein (1974); Hanushek, et al (1974)).

Investigations of contextual effects (e.g., the effects of school environment on performance) have been particularly prone to the problems of model misspecification. Hauser (1970, 1971, 1972a, 1974) has demonstrated that so-called context effects virtually disappear once all relevant individual-level variables have been included in the model.

Appropriate Cross-Level and Sub-Level Designs

There are situations which call for multilevel analyses of data. We cite below three studies in this vein.

Group Anchored Verses Sample Anchored Measures. Lewy (1972) suggests that the appropriate type of correlation coefficient is a function of both questions one asks and the type of variables one examines. He points out that if measures are "group-anchored" (relative referents such as teacher grades and student's self-appraisal) as opposed to "sample-anchored" (absolute referent such as standardized achievement measures), then pooled within-group correlations, rather than total individual-level correlations, convey the right information. Furthermore, Lewy demonstrates how the relative magnitudes and signs of between-group correlations, pooled within-group correlations and total individual correlations vary according to type of measure (group-anchored or sample-anchored). The data he presents are provocative and suggest that multilevel analyses are necessary for school effects data.

Within-College and Between-College Analyses. Rock et al (1972) attempted to find groups of colleges that are equally effective, using within college variation to estimate interactions between student aptitude and college effects. They (a) calculated within-college regression lines, (b) clustered colleges on the parameters of the within-college regressions, (c) generated discriminant functions to check for statistical distinctions among clusters of colleges on the basis of the within-college parameters, and (d) identified the descriptive measures of the colleges that successfully discriminate among the clusters. Thus, Rock and his associates utilized both within-school and between-school analyses to achieve their objectives.

Analysis of Hierarchical Data. Kiesling and Wiley (1973) argue that school-level indices such as average daily attendance do not convey independent information for each pupil within the school and thus should not be included in between-student analyses. If there are only say, 40 schools, then there are 39 degrees of freedom for school, no matter how many students are involved. They advocate (a) performing within-school regressions, (b) aggregating the information to the level of the school, and (c) entering the aggregated within-school model as a variable in a between-school analysis in order to minimize bias in estimating parameters at the school level. Kiesling and Wiley demonstrate their techniques with data from the Coleman study, and they indeed improve the estimation of the effects of school inputs in this fashion.

The three studies discussed above reflect new and, perhaps improved, directions for the analyses of effects of schooling. Each analysis is conditioned on the questions asked and the process by which variables are generated and can logically affect one another. They also demonstrate once again the complexity of choosing the appropriate units and in doing so, aptly summarize our conclusions.

FOOTNOTES

[1]This paper is an outgrowth of research supported by the National Institute of Education (Grant No. NIE-C-74-0123 to Vasquez Associates, Ltd.). Lee J. Cronbach, Edward L. Feige, Michael T. Hannan, Robert M. Hauser, Thomas R. Knapp, Harry Lütjohan, Carlyle E. Maw, Ingram Olkin, and David E. Wiley have all made contributions to the ideas expressed herein through their comments and suggestions over the past 3 years. Michael Baenen, James Knoop, and Jan Shanahan assisted with the data analysis for Appendices A and B and Margie Mika prepared readable copies of the manuscript. The errors and misrepresentations that remain are solely the responsibility of the author.

[2]This list does not even begin to represent the work on related topics such as the study of contextual effects. See, e.g., Hauser, 1972, 1974.

[3]We are trying to distinguish between data such as provided by the U.S. Census and the data from the IEA survey.

[4]See Feige and Watts (1972) for an example involving Federal Reserve data.

[5]Boruch (1971, 1972) mentions the procedure in his writings on conducting research with confidential data.

[6]The author wishes to thank Torsten Husen and Roy W. Phillipps for their permission to reproduce the data from the Technical Report of the IEA study.

APPENDIX A: Empirical examples of the estimation of standardized regression coefficients in the single regressor case -- A discussion of Hannan-Burstein bias prediction, and Feige-Watts techniques.

This appendix performs two functions. First, it provides two empirical examples that are more comprehensive than those published in Hannan and Burstein (1974) and Burstein (1974a). All variables have been standardized so that we are estimating correlation coefficients as well as regression coefficients in this single regressor case.

Second, the examples also represent the first appearance in the educational literature of techniques developed by the economists Feige and Watts (1972). Their techniques provide a statistic for the discrepancy between grouped and ungrouped estimates of regression parameters. In the simple models presented here, we are able to consider utility of the Feige-Watts formulation relative to the prediction of bias from the Hannan-Burstein approach.

Table A.1 includes a short description of each grouping variable. The primary variables in the two examples are self appraisal of academic abilities (SRAA), a weighted composite of ten items asking the person to rate himself in various academic skill areas; total score from an achievement test battery given during college orientation (ACH); and total score on the Scholastic Aptitude Test (SAT). In each example we wish to estimate β_{YX} from the regression

$$Y = \beta_{YY}X + u \quad .$$

Tables A.2 and A.3 contain estimates of important parameters from the structural equations for the standardized regressions of SRAA on ACH and ACH and SAT and a grouping variable Z: (See Hannan and Burstein (1974); and Burstein (1974a) for details on the development of the approach.)

Ungrouped: $$Y = \beta_{YX\cdot Z}X + \beta_{YZ\cdot X}Z + w$$

$$X = \beta_{XZ}Z + v$$

Grouped: $$\bar{Y} = \beta_{YX\cdot Z}\bar{X} + \beta_{YZ\cdot X}\bar{Z} + \bar{w}$$

$$\bar{X} = \beta_{XZ}\bar{Z} + \bar{v}$$

In Burstein (1974a)and in the present discussion, we categorized the grouping variables (Z) according to the following procedure:

I. Z directly related to both Y and X -- $|\hat{\beta}_{YZ\cdot X}|>3SE(\hat{\beta}_{YZ\cdot X});|\hat{\beta}_{XZ}|>3SE(\hat{\beta}_{XZ})$.

II. Z directly related to Y but not to X --
$|\hat{\beta}_{YZ\cdot X}|>3SE(\hat{\beta}_{YZ\cdot X});|\hat{\beta}_{XZ}|\leq 3SE(\hat{\beta}_{XZ})$.

III. Z directly related to X but not to Y --
$|\hat{\beta}_{YZ\cdot X}|\leq 3SE(\hat{\beta}_{YZ\cdot X});|\hat{\beta}_{XZ}|>3SE(\hat{\beta}_{XZ})$.

IV. Z not directly related to either Y or X --
$|\hat{\beta}_{YZ\cdot X}|\leq 3SE(\hat{\beta}_{YZ\cdot X});|\hat{\beta}_{XZ}|\leq 3SE(\hat{\beta}_{XZ})$.

Our expectation is that Category III grouping variables yield the best estimates (in terms of small bias and mean squared error) because they most nearly parallel grouping on the regressor (which is known [Prais and Aitchinson (1954), Cramer (1964)] to yield the best estimates of all possible grouping procedures.) Category IV grouping should yield estimates that are unbiased but are relatively inefficient. This occurs because grouping on a Category IV variable which forms m groups yields essentially the same results as basing one's estimate on a random sample of m observations drawn from the original N. (See Cramer (1964) and Feige and Watts (1972) for details). We have yet to see a Category II grouping variable but expect it to behave more like Category IV grouping than any other category. Category I variables should yield relatively poor estimates which tend to be inefficient as well as biased.

Tables A.4 and A.5 contain the estimates from grouped data of the standardized regression coefficients and their standard errors for the two models plus estimates of the predicted bias based on the formula presented in both Hannan-Burstein (1974, essentially equation (17) on page 386 for the standardized case) and Burstein (1974a, page 27, again modified slightly) and at the bottom of each table.[1] Feige-Watts F values are also included in the tables and will be discussed later on. A detailed discussion of these tables appears elsewhere (Burstein, (1975)) but a few observations can be made.

1. With only one exception (CLIMP in the regression of SRAA on ACH), Category III variables yield estimates with smaller bias than any other variables except ID2, which formed at least ten times as many groups.
2. The standard errors for all estimates are relatively large which is not surprising given the limited number of groups in most cases.[2]
3. The standard errors for Category III grouping are generally smaller than those from grouping on a Category I variable which forms a comparable number of groups.
4. The truly poor grouping methods (e.g., ANTDEG and QCJOB in the regression of SRAA on ACH) are clearly identified and their huge biases predicted with accuracy. The 8 (out of 17) variables with largest predicted biases had the largest observed biases for the regression of SRAA on ACH; the 6 variables with the largest predicted and observed biases from the regression of ACH on SAT were also the same.
5. The results of predicting the bias for Category III and IV grouping were mixed. There are sign differences between

---

[1] Grouping on the regressor and regressand have been placed first in their respective categories (III and I) to indicate their special significance.

[2] The computer procedure used to find the between-group regressions weights each group mean by the number of observations in the group and thereby bases the standard errors of the coefficients on the entire sample of 2676 observations. We have rescaled the standard errors from the grouped data by multiplying by $\sqrt{(N-1/(m-1)}$ to reflect the actual number of observations on which each coefficient is based.

predicted and observed bias for some variables (PARINC in the ACH-on-SAT regression, ID1 in the SRAA-on-ACH (regression); some cases of underprediction by as much as .1 (ID1 in the ACH-on-SAT regression, CLIMP in the SRAA-on-ACH regression); and the predicted bias for PARINC in the SRAA-on-ACH regression is .1 larger than the observed. However, overall, 7 out of 10 of the groupings from both models (5 from each) with smallest observed bias were among the 10 with smallest predicted bias. Only 2 of the 14 variables with predicted bias less than $|.1|$ had observed bias greater than $|.1|$ (ID1 in the ACH-on-SAT regression and CLIMP in the SRAA-on-ACH regression).

The moderate success of our attempts to predict bias in these simple cases is encouraging, given the small number of groups formed by most variables and problems such as nonlinearity of relations, skewness and other factors affecting the grouping characteristics. By use of compositing techniques first described in Burstein (1974a) and discussed in Appendix B, we can improve the accuracy of our estimates of individual-level parameters.

Feige and Watts Technique

Feige and Watts (1972) developed a measure of the divergence between grouped and ungrouped estimators, $\underset{\sim}{\beta}$ and $\underset{\sim}{b}$, in the multivariate case. They attributed this divergence to three sources -- (i) specification bias, (ii) bias introduced by a grouping that is not independent of the disturbances from the structural model, and (iii) sampling error induced by the loss of information in grouping. We summarize the Feige-Watts analysis below.

We are interested in estimating $\underset{\sim}{\beta}$ from the model

$$\underset{\sim}{Y} = \underset{\approx}{X}\underset{\sim}{\beta} + \underset{\sim}{u} \quad .$$

The least-squares estimators of the regression parameters from raw data and their variance-covariance matrix are given by

$$\underset{\sim}{b} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$$

and

$$V(\underset{\sim}{b}) = \sigma_u^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \quad .$$

In order to generate grouped data, a m x N grouping matrix $\underset{\sim}{G}$ is introduced which transforms the raw data to a set of m rows. The ith rows of the transformed matrices contain the mean values of the variables for the ith group; i.e., the matrix $(\underset{\sim}{Y}, \underset{\sim}{X})$ is replaced by

$$(\bar{\underset{\sim}{Y}},\bar{\underset{\sim}{X}}) = (\underset{\sim}{G}\underset{\sim}{Y},\underset{\sim}{G}\underset{\sim}{X}) \quad .$$

Let $\underset{\sim}{H}$ be a N x N matrix which produces the same grouping as $\underset{\sim}{G}$, but replicates the mean rows to accomplish the weighting that is necessary for unequal group sizes. $\underset{\sim}{H}$ is related to $\underset{\sim}{G}$ by:

$$\underset{\sim}{H} = \underset{\sim}{G}'(\underset{\sim}{G}\underset{\sim}{G}')^{-1}\underset{\sim}{G}.$$

Using the $\underset{\sim}{H}$ transformation, the estimates of $\underset{\sim}{\beta}$ and their covariance matrix from grouped data can be written

$$\underset{\sim}{B} = (\underset{\sim}{X}'\underset{\sim}{H}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{H}\underset{\sim}{Y}$$

and

$$V(\underset{\sim}{B}) = \sigma^2(\underset{\sim}{X}'\underset{\sim}{H}\underset{\sim}{X})^{-1} \quad .$$

The divergence of grouped and ungrouped estimates of $\underset{\sim}{\beta}$,

$$\Delta(\underset{\sim}{H}) = \underset{\sim}{b} - \underset{\sim}{B} \ ,$$

has a zero mean and variance-covariance matrix equal to

$$\mathrm{cov}[\Delta(\underset{\sim}{H})] = \sigma^2\{(\underset{\sim}{X}'\underset{\sim}{H}\underset{\sim}{X})^{-1} - (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\} \quad .$$

Let $\bar{\underset{\sim}{e}} = \bar{\underset{\sim}{Y}} - \bar{\underset{\sim}{X}}\underset{\sim}{B}$ so that $\bar{\underset{\sim}{e}}'\bar{\underset{\sim}{e}}$ is the sum of squared residuals from the between-groups regression. Assume additionally that the disturbances $\underset{\sim}{u}$ are normally distributed. Then the quadratic forms

$$Q_1 = \frac{\Delta'(\underset{\sim}{H})[(\underset{\sim}{X}'\underset{\sim}{H}\underset{\sim}{X})^{-1} - (\underset{\sim}{X}'\underset{\sim}{X})^{-1}]^{-1}\Delta(\underset{\sim}{H})}{\sigma_u^2}$$

and

$$Q_2 = \frac{\bar{\underset{\sim}{e}}'\bar{\underset{\sim}{e}}}{\sigma_u^2}$$

are χ^2 variables with k and m-k degrees of freedom, respectively.

If the model is correctly specified and H and u are independent,

$$F = \frac{(Q_1/k)}{(Q_2/[m-k])}$$

is distributed as an F statistic with k and m-k degrees of freedom. Values of F beyond the critical values of the F-distribution indicate differences between estimators that could not be attributed to sampling error. Hence good grouping methods yield small F values.

To illustrate their findings, Feige and Watts examined 20 regression equations generated from income and dividend information provided by 5393 banks to the Federal Reserve System. The seven grouping rules they used included a random procedure and geographic and financial asset indices. There were also 3 levels of aggregation -- slight (3 observations per group), moderate (30 observations) and drastic (100 observations). Thus 21 grouping methods were possible for each equation although the article only discussed a few.

Certain of the Feige-Watts equations were quite sensitive to the choice of grouping method. The reported F values ranged from .02 to 84.96. All the F values were significant at the .05 level for one equation, while grouping produced no significant F tests for other equations.

In every case, slight aggregation was superior to other levels.

Thus, a large number of groups again proved to be desirable.

The models in our example are much less complex than the ones Feige and Watts considered. Because we have only a single regressor and the variables are standardized, the formula for Feige-Watts measure of divergence simplifies to

$$F = \frac{\Delta^2 \quad (\hat{\sigma}(\Delta))}{\tilde{\sigma}\ (\overline{res})/m-1}$$

where $\Delta = \beta_{\bar{Y}\bar{X}} - b_{YX}$

$$\hat{\sigma}(\Delta) = \left[\frac{1}{\sigma(\bar{X})} - \frac{1}{\sigma(X)}\right]^{-1} = (N-1)\ \frac{\left(1 - \hat{\sigma}^2_{\bar{X}}\right)}{\hat{\sigma}^2_{X}}.$$

and $(\overline{res})$ is residual sum of squares from the aggregate regression.

As can be seen from Tables A.4 and A.5, there are, in general, large F statistics for what we have called Category I grouping variables and small F statistics for Category III and IV grouping. Grouping on the regressand provides the largest F statistic and grouping on the regressor, the smallest. In fact, 11 of the 12 variables with F values less than 1.0 result in observed biases smaller than .1.

The Hannan-Burstein bias predictions and the Feige-Watts statistics both show signs of promise for providing guidance in choosing the optimal grouping methods. But, as our example demonstrates, the results from these procedures do not always conform to our expectations. Until we can identify only good grouping methods and eliminate all poor ones, we will have to continue to improve our understanding of the grouping process.

Table A.1 Information on grouping variables.

Variable Identification	Description	Number of Groups After Aggregation
ID2	Last 2 digits of student identification	100
ID1	Last digit of student identification	10
HSGPA2	High school's report of student's grade point average on a 4-point scale (highest 2 digits)	23
SAT2	Highest 2 digits of Total score from the Scholastic Aptitude Test	13
ACH2	Highest 2 digits of Total score from the Achievement Battery	10
PARINC	Student's best estimate of 1970 parental income before taxes	10
REPGPA	Student's report of average grade in secondary school	7
POPED	Student's report of highest level of formal education obtained by his father	6
SRAA2	Highest digit and sign of composite academic self-opinion	5
ANTHIDEG	Student's anticipated highest academic degree	5
HSMATH	Student's report of number of semesters of high school mathematics	5
HSPHYS	Student's report of number of semesters of high school physical sciences	5
NOBOOK	Student's report of number of books in the home	5
PARASP	"What is the highest level of education that your parents hope you will complete?"	5

Table A.1(Continued). Information on grouping variables.

CLIMP	"My grades are markedly better in courses that I see I will need later."	4
COLEFF	"I often wonder if four years of college will really be worth the effort."	4
QCJOB	"I often wish that I were offered a good job now so I wouldn't have to spend four years in college."	4

Table A.2 Estimates of parameters relating ACH(X) and SRAA(Y) to possible grouping variables (Z)[a]

VARIABLE NAME	GROUP SIZE (m)	$\hat{\beta}_{YZ\cdot X}$	PARAMETER ESTIMATES $\hat{\beta}_{XZ}$	$\hat{\beta}_{YZ}$	$\hat{\sigma}_{\bar{X}}$
ID2	100	.008	.020	.019	.189
ID1	10	-.011	-.042	-.033	.078
HSGPA2	23	.123	.535	.370	.552
SAT2	13	.406	.827	.566	.831
ACH2	10	.070	.983	.522	.984
PARINC	10	.028	.070	.064	.122
REPGPA	7	-.258	-.490	-.455	.510
POPED	6	.073	.139	.145	.150
SRAA2	5	.819	.476	.885	.481
ANTDEG	5	.186	.156	.264	.159
HSMATH	5	-.066	.479	.202	.489
HSPHYS	5	.046	.318	.209	.365
NOBOOK	5	.122	.146	.196	.148
PARASP	5	.138	.066	.186	.077
CLIMP	4	.003	.147	.074	.163
COLEFF	4	.121	.134	.189	.144
QCJOB	4	.145	.105	.199	.113

[a]All variables have been standardized prior to grouping so that $\sigma_Y = \sigma_X = \sigma_Z = 1$, $\beta_{XZ} = \rho_{XZ}$, and $\beta_{YZ} = \rho_{YZ}$.

Table A.3 Estimates of parameters relating SAT(X) and ACH(Y) to possible grouping variables (Z)[a]

VARIABLE NAME	GROUP SIZE (m)	PARAMETER ESTIMATES $\hat{\beta}_{YZ\cdot X}$	$\hat{\beta}_{XZ}$	$\hat{\beta}_{YZ}$	$\hat{\sigma}_{\bar{X}}$
ID2	100	.014	.008	.020	.186
ID1	10	-.003	-.046	-.042	.069
HSGPA2	23	.164	.488	.535	.517
SAT2	13	-.042	.987	.828	.989
ACH2	10	.916	.827	.983	.835
PARINC	10	.006	.076	.070	.146
REPGPA	7	-.124	-.468	-.490	.498
POPED	6	.007	.157	.139	.169
SRAA2	5	.054	.520	.476	.531
ANTDEG	5	.039	.140	.108	.141
HSMATH	5	.214	.346	.480	.349
HSPHYS	5	.109	.257	.318	.294
NOBOOK	5	-.025	.203	.146	.204
PARASP	5	-.007	.087	.066	.101
CLIMP	4	.009	.165	.147	.185
COLEFF	4	.039	.114	.134	.134
QCJOB	4	.007	.118	.106	.123

[a]All variables have been standardized prior to grouping so that $\sigma_Y = \sigma_X = \sigma_Z = 1$, $\beta_{XZ} = \rho_{XZ}$, and $\beta_{YZ} = \rho_{YZ}$.

Table A.4 Estimates from grouped data of the standardized coefficients from the regression of SRAA on ACH, and bias prediction using Hannan-Burstein and Feige-Watts procedures.

VARIABLE NAME	NUMBER OF GROUPS (m)	$B_{\bar{Y}\bar{Z}}$ [a]	$SE(B_{\bar{Y}\bar{X}})$ [a]	OBSERVED BIAS Δ [b]	HANNAN-BURSTEIN PREDICTED BIAS $\hat{\theta}$ [c]	FEIGE-WATTS F [d]
CATEGORY IV						
ID2	100	.558	.0739	.029	.004	.160
ID1	10	.442	.1831	-.087	.075	.228
CATEGORY III						
ACH2	10	.531	.0615	.002	.002	.049
PARINC	10	.558	.1314	.029	.129	.051
HSPHYS	5	.571	.0915	.043	.095	.252
CLIMP	4	.717	.3971	.187	.016	.230
CATEGORY I						
SRAA2	5	1.853	.0631	1.324	1.295	571.662**
HSMATH	5	.414	.0248	-.115	-.100	27.823**
SAT2	13	.671	.0670	.142	.150	14.517**
HSGPA2	23	.702	.0287	.173	.150	52.260**
POPED	6	.911	.1626	.382	.440	5.635
REPGPA	7	.917	.0617	.388	.360	53.256**
NOBOOK	5	1.334	.1133	.805	.800	51.760**
COLEFF	4	1.461	.1160	.932	.765	49.319**
ANTDEG	5	1.631	.2680	1.102	1.117	16.280**
QCJOB	4	1.853	.3533	1.324	1.188	14.227*
PARASP	5	1.946	.7339	1.417	1.519	3.752

[a] Estimates from ingrouped data: $b_{YX} = .529$; $SE(b_{YX}) = .0032$.

[b] Observed Bias = $\Delta = B_{\bar{Y}\bar{X}} - b_{YX}$

[c] Predicted Bias = $\hat{\theta} = \hat{\beta}_{YZ \cdot X}\hat{\beta}_{XZ} \dfrac{(1 - \hat{\sigma}_{\bar{X}}^2)}{\hat{\sigma}_{\bar{X}}^2}$.

[d] Feige-Watts F = $\dfrac{\Delta^2(SS(\Delta))}{SS(\overline{res})/m-1}$

* Exceeds the 95 percent critical value for F.

** Exceeds the 99 percent critical value for F.

Table A.5 Estimates from grouped data of the standardized coefficients from the regression of ACH on SAT and bias prediction using Hannan-Burstein and Feige-Watts procedures.

VARIABLE NAME	NUMBER OF GROUPS (m)	$B_{\bar{Y}\bar{Z}}$ [a]	$SE(B_{\bar{Y}\bar{X}})$ [a]	OBSERVED BIAS Δ [b]	HANNAN-BURSTEIN PREDICTED BIAS $\hat{\theta}$ [c]	FEIGE-WATTS F [d]
CATEGORY IV						
ID2	100	.832	.0059	-.007	.003	.015
ID1	10	1.053	.2168	.214	.029	2.469
CATEGORY III						
SAT2	13	.838	.0190	-.001	-.001	.000
PARINC	10	.817	.0598	-.022	.021	.143
CLIMP	4	.876	.0388	.036	.042	.904
POPED	6	.877	.0685	.039	.038	.331
SRAA2	5	.899	.0543	.060	.007	.177
QCJOB	4	.912	.0261	.073	.054	7.982
PARASP	5	.744	.0903	-.095	-.059	1.118
NOBOOK	5	.718	.0372	-.121	-.174	11.096**
CATEGORY I						
ACH2	10	1.168	.0541	.329	.329	197.730**
REPGPA	7	1.019	.0418	.180	.176	25.398**
COLEFF	4	1.054	.1169	.214	.241	3.438
HSGPA2	23	1.057	.0329	.218	.219	61.750**
ANTDEG	5	1.120	.0607	.281	.271	21.759**
HSPHYS	5	1.237	.0422	.398	.295	98.739**
HSMATH	5	1.396	.0478	.557	.531	149.284**

[a] Estimates from ungrouped data: $b_{YX} = .839$; $SE(b_{YX}) = .0105$

[b] Observed Bias $= \Delta = B_{\bar{Y}\bar{X}} - b_{YX}$

[c] Predicted Bias $= \hat{\theta} = \hat{\beta}_{YZ\cdot X}\hat{\beta}_{XZ} \dfrac{(1 - \hat{\sigma}_{\bar{X}}^{2})}{\hat{\sigma}_{\bar{X}}^{2}}$

[d] Feige-Watts $F = \dfrac{\Delta^2(SS(\Delta))}{SS(\overline{res})/m-1}$

* Exceeds the 95 percent critical value for F.

** Exceeds the 99 percent critical value for F.

APPENDIX B: Estimating individual-level relations from grouped data when some data on individuals has been collected anonymously.

The data described in Appendix A are also used for the examples in this appendix. However, here we assume that information on one of the primary variables is collected anonymously while information on the other is identifiable (Context (D)). We are currently investigating various possibilities for circumventing this lack of information through aggregation techniques. Some preliminary findings are described below.

Predicted Bias when σ_{XY} is Unknown -- π. The formula for predicting bias due to grouping when the variables are standardized at the individual level is

$$\theta = \beta_{YZ \cdot X}\beta_{XZ} \left(\frac{1 - \sigma_{\bar{X}}^2}{\sigma_{\bar{X}}^2}\right) ,$$

where $\beta_{YZ \cdot X}$ and β_{XZ} are standardized regression coefficients and $\sigma_{\bar{X}}^2$ is the between group variance of the regressor.

In order to estimate $\beta_{YZ \cdot X}$, we need to know σ_{XY}, the covariance between regressor and regressor. But when data either X or Y has been collected anonymously and X and Y scores cannot be matched at the individual level, we are unable to estimate σ_{XY}. Thus we are unable to estimate θ in this context.

A promising alternative is to utilize the information conveyed by the standardized coefficients from the regressions of Y on Z (β_{YZ}) and X and Z (β_{XZ})[1] to identify good grouping methods. If $\beta_{YZ} > \beta_{YZ \cdot X}$, then an upper bound[2] for the predicted bias is given by

$$\pi = \beta_{YZ}\beta_{\bar{X}\bar{Z}} \left(\frac{1 - \sigma_{\bar{X}}^2}{\sigma_{\bar{X}}^2}\right)$$

$$= \frac{\beta_{YZ}}{\beta_{YZ \cdot X}} \theta$$

In Tables B.1 and B.2 information from Tables A.4 and A.5 has been reproduced with the following modifications:

(a) the Feige-Watts values have been deleted and

(b) π values based on the information contained in Tables A.2 and A.3 have been included.

Tables B.1 and B.2 here

The categorization scheme introduced by Burstein (1974a) is not as useful for distinguishing among grouping variables with high and low predicted π's as it was with the predicted θ's. The estimates of π from grouping variables tend to be inflated relative to the co esponding estimate of θ when β_{YZ} is small and $\beta_{XZ}>\beta_{YZ}$ (e.g., NOBOOK and POPED in the regression of ACH on SAT). π estimates are depressed relative to the corresponding θ's when β_{XZ} and β_{YZ} are both moderate to large in magnitude and $\beta_{YZ}>\beta_{XZ}$ (e.g., HSGPA2 in the regression of ACH on SAT).

Thus the rank ordering of grouping variables according to their θ and π values will differ. However, the problem of identifying good grouping methods is not completely hampered by this inconsistency if the investigator utilizes other common sense guidelines.

First, he should group only variables for which $\beta_{YZ} > \beta_{XZ}$. This constraint would again eliminate most Category I variables.

Second, any grouping variable should be eliminated for which $B_{\bar{Y}\bar{X}}$ falls outside the bounds of possible values of β_{YX}. In this case, grouping variables which yield values of $B_{\bar{Y}\bar{X}}$ greater than 1 should be eliminated since we are estimating a standardized coefficient.

Table B.1 Estimates from grouped data of the standardized coefficients from the regression of SRAA on ACH, and predicted bias based on θ and π.

VARIABLE NAME	NUMBER OF GROUPS (m)	$B^{a}_{\bar{Y}\bar{X}}$	$SE(B_{\bar{Y}\bar{X}})^{a}$	OBSERVED BIAS Δ	PREDICT BIAS USING $\hat{\theta}^{c}$	$\hat{\pi}^{d}$
CATEGORY IV						
ID2	100	.558	.0739	.029	.004	.040
ID1	10	.442	.1831	-.087	.075	.225
CATEGORY III						
ACH2	10	.531	.0615	.002	.002	.042
PARINC	10	.558	.1314	.029	.129	.295
HSPHYS	5	.571	.0915	.043	.095	.433
CLIMP	4	.717	.3971	.187	.016	.401
CATEGORY I						
SRAA2	5	1.853	.0631	1.324	1.395	1.507
HSMATH	5	.414	.0248	-.115	-.100	.307
SAT2	13	.671	.0670	.142	.150	.210
HSGPA2	23	.702	.0287	.173	.150	.451
POPED	6	.911	.1626	.382	.440	.874
REPGPA	7	.917	.0617	.388	.360	.635
NOBOOK	5	1.334	.1133	.805	.800	1.285
COLEFF	4	1.461	.1160	.932	.765	1.194
ANTDEG	5	1.631	.2680	1.102	1.117	1.586
QCJOB	4	1.853	.3533	1.324	1.188	1.630
PARASP	5	1.946	.7339	1.417	1.519	2.048

[a] Estimates from ungrouped data: $b_{YX} = .529$; $SE(b_{YX}) = .0032$.

[b] Observed Bias $= \Delta = B_{\bar{Y}\bar{X}} - b_{YX}$

[c] Predicted Bias $= \hat{\theta} = \hat{\beta}_{YZ \cdot X}\hat{\beta}_{XZ} \left(\frac{1 - \hat{\sigma}_{\bar{X}}^2}{\hat{\sigma}_{\bar{X}}^2}\right)$

[d] $\hat{\pi} = \hat{\beta}_{YZ}\hat{\beta}_{XZ} \left(\frac{1 - \hat{\sigma}_{\bar{X}}^2}{\hat{\sigma}_{\bar{X}}^2}\right) = \frac{\hat{\beta}_{YZ}}{\hat{\beta}_{YZ \cdot X}} \hat{\theta}$

Table B.2 Estimates from grouped data of the standardized coefficients from the regression of ACH on SAT and bias prediction based on θ and π.

VARIABLE NAME	NUMBER OF GROUPS (m)	$B^{a}_{\overline{YX}}$	$SE(B_{\overline{YX}})^{a}$	OBSERVED BIAS Δ	PREDICT BIAS USING $\hat{\theta}^{c}$	$\hat{\pi}^{d}$
CATEGORY IV						
ID2	100	.832	.0059	-.007	.003	.004
ID1	10	1.053	.2168	.214	.029	.406
CATEGORY III						
SAT2	13	.838	.0190	-.001	-.001	.020
PARINC	10	.817	.0598	-.022	.021	.252
CLIMP	4	.876	.0388	.036	.042	.686
POPED	6	.877	.0685	.039	.038	.751
SRAA2	5	.899	.0543	.060	.007	.067
QCJOB	4	.912	.0261	.073	.054	.818
PARASP	5	.744	.0903	-.095	-.059	.533
NOBOOK	5	.718	.0372	-.121	-.174	1.016
CATEGORY I						
ACH2	10	1.168	.0541	.329	.329	.352
REPCPA	7	1.019	.0418	.180	.176	.695
COLEFF	4	1.054	.1169	.214	.241	.828
HSGPA2	23	1.057	.0329	.218	.219	.095
ANTDEG	5	1.120	.0607	.281	.271	.751
HSPHYS	5	1.237	.0422	.398	.295	.858
HSMATH	5	1.396	.0478	.557	.531	1.189

[a]Estimates from ungrouped data: $b_{YX} = .529$; $SE(b_{YX}) = .0032$.

[b]Observed Bias $= \Delta = B_{\overline{YX}} - b_{YX}$

[c]Predicted Bias $= \hat{\theta} = \hat{\beta}_{YZ\cdot X}\hat{\beta}_{XZ} \dfrac{(1 - \hat{\sigma}_{\bar{X}}^2)}{\hat{\sigma}_{\bar{X}}^2}$

[d]$\hat{\pi} = \hat{\beta}_{YZ}\hat{\beta}_{XZ} \left(\dfrac{1 - \hat{\sigma}_{\bar{X}}^2}{\hat{\sigma}_{\bar{X}}^2}\right) = \dfrac{\hat{\beta}_{YZ}}{\hat{\beta}_{YZ\cdot X}}\hat{\theta}$

Finally, grouping variables for which the predicted bias is very large can be dropped from further consideration.

This last guideline is the most controversial and ambiguous as it is not obvious what a very large π" will be in any particular study. Clearly, the predicted π (1.016) for NOBOOK in the ACH-on-SAT regression is very large, but are the π's for POPED (.751) and CLIMP (.686) also very large? Not necessarily; it all depends on the magnitude of the estimates of β_{YX} from other grouping variables and on the magnitude of π for other potential Z's.

In any case, it is reassuring that in both tables, the variables that have lower π values and also meet the three common sense guidelines provide reasonably accurate estimates of β_{YX}. In the next section we describe an approach that combines variables with smallest predicted bias and thereby affords greater confidence in estimates from grouped data.

Composite Estimates from Multiple Grouping Variables. The above findings suggest that even in Context (D), an investigator can distinguish those grouping characteristics which lead to reasonably accurate estimates from those providing extremely misleading ones. Once this separation has been accomplished, the investigator can choose the characteristic with the smallest predicted bias. Better yet, he can use the available information about each characteristic and its expected bias to form a weighted composite of good grouped estimates. For example, grouped estimates can be weighted in an inverse proportion to their predicted bias. The standard errors of the grouped estimates or the number of groups formed (m) can also be used to give additional weight to the potentially more stable estimates . In Context (D) where σ_{YX} is unknown, we cannot group on the regressor, and we are trying to generate estimates that are better than those from random grouping (Category IV). Our compositing procedure works as follows:

(1) Identify the 3 grouping variables (other than the regressor or Category IV variables) with smallest $\hat{\pi}$ values, excluding those variables for which $B_{\overline{X}\overline{Y}} \geq 1.0$.

(2) Find the Fisher Z-transformation for the $B_{\overline{Y}\overline{X}}$ of each of the 3 variables.

(3) Weight the Fisher-Z values:

(a) in inverse proportion to the predicted biases

$$\hat{\pi}* = (\sum_{i=1}^{3} \hat{\pi}_i - \hat{\pi}_i)$$ where i identifies a specific grouping variable;

(b) by the number of groups formed (m);

(c) in inverse proportion to the squared standard error of the grouped estimate:

$$(SE(B_{\overline{Y}\overline{X}}))^{B*} = (\sum_{i=1}^{3} (SE(B_{\overline{Y}\overline{X}})^2{}_i) - (SE(B_{\overline{Y}\overline{X}}))^2_i$$

(4) Find the weighted average of Fisher-Z values in each case:

(5) Transform the average Fisher-Z back to correlation units to find the composite estimate of $B_{\overline{Y}\overline{X}}$.

There are other possible weighting methods and any number of weighting methods can be combined to generate a new weighting scheme. The ones included here are intended only to illustrate the technique.

In Table B.3, we present composite estimates for the standardized coefficient from 3 regressions -- (A) SRAA on ACH, (B) ACH on SAT, (C) SRAA on SAT. Information on the grouping variables that contribute to the composites is also provided.

Table B.3 here

The results of the compositing process are satisfactory. In each example at least one of the composite estimates is more accurate than any Category III or Category IV variable with the exception of grouping on the regressor. The practical utility of the composite estimates is high

Table B.3 Weighted composites from grouped estimates of standardized regression coefficients

A. Regression of SRAA on ACH -- b_{YX} = .529

Grouping Variable	No. of Groups (m)	$B_{\overline{YX}}$	$SE(B_{\overline{YX}})$	Predicted Bias (π)
PARINC	10	.558	.1314	.295
HSMATH	5	.414	.0248	.307
SAT2	13	.671	.0670	.210

Estimates from Weighted composites

Weights Determined by	Estimate
π*	.564
m	.594
$(SE(B_{\overline{YX}}))$*	.545

B. Regression of ACH on SAT -- b_{YX} = .839

Grouping Variable	No. of Groups (m)	$B_{\overline{YX}}$	$SE(B_{\overline{YX}})$	Predicted Bias (π)
PARINC	10	.817	.0598	.252
SRAA2	5	.899	.0543	.067
PARASP	5	.744	.0903	.533

Estimates from Weighted composites

Weights Determined by	Estimate
π*	.852
m	.828
$(SE(B_{\overline{YX}}))^{2*}$	.844

Table B.3 (C) cont.

C. Regression of SRAA on SAT -- $b_{YX} = .574$

Grouping Variable	No. of Groups	$B_{\overline{Y}\overline{X}}$	$SE(B_{\overline{Y}\overline{X}})$	Predicted Bias ($\hat{\pi}$)
ACH2	10	.651	.0750	.575
CLIMP	4	.672	.3960	.637
PARINC	10	.434	.1316	.237

Estimates from Weighted composites

Weights Determined by	Estimate
π*	.576
m	.574
$(SE(B_{\overline{Y}\overline{X}}))^{2*}$	.564

in every case since the worst of the composite estimates (.594, based on the SRAA-on-ACH regression where b_{YX} = .529) deviates by only 12% from the ungrouped value. This implies a .12 standard deviation over prediction from this compositing procedure.

Much remains to be done before we can prescribe uniformly powerful methods of describing how many variables to include in the composite and what method should be used to weight the estimates. However, it is obvious that the compositing process has merit especially when more direct approaches to choosing the best grouping variable are not possible as in Context (D).

Reconstructing a Correlation Matrix from Grouped Data. We are also exploring the feasibility of accurately reconstructing individual-level correlation matrices from estimates based on grouped observations. The procedures we have examined require the standardization of all primary variables before grouping, and knowledge of the zero-order correlation coefficients relating each grouping variable (Z) to the primary variables (here designated X_h). These correlations are presented in Table B.4.

Table B.4 here

For any two primary variables X_1 and X_2, and each grouping variable Z, we regressed $\bar{X}_1$ on $\bar{X}_2$ (weighted group means) when r_{X_2Z} is moderate to large and $r_{X_2Z} > r_{X_1Z}$. When r_{X_1Z} is moderate to large and $r_{X_1Z} > r_{X_2Z}$, we regressed $\bar{X}_Z$ and $\bar{X}_1$. This procedure yields a pool of $B_{\bar{X}_1\bar{X}_2}$ and $B_{\bar{X}_Z\bar{X}_1}$ values that can be used to estimate $r_{X_1X_2}$.

By examining the estimates generated in the above fashion, we were successful in reconstructing a correlation matrix. In Table B.5, the individual-level correlations coefficients among five primary variables are presented in the upper triangular portion of the matrix. The best estimates from grouping on a variable other than the regressor or random

Table B.4 Zero-Order correlations relating potential grouping variables to primary variables.

Grouping Variable	CORRELATION COEFFICIENTS Primary Variables SRAA	ACH	SAT	HSGPA	HSPHYS
ANTDEG	.264	.156	.140	.046	.189
CLIMP	.074	.147	.165	.139	.031
COLEFF	.189	.134	.114	.071	.090
HSMATH	.202	.479	.346	.248	.358
HSPHYS	.209	.318	.257	.132	---
NOBOOK	.196	.146	.203	.030	.021
PARASP	.172	.066	.087	.001	.109
PARINC	.064	.070	.076	-.101	-.009
POPED	.145	.139	.157	-.010	.007
QCJOB	.199	.106	.118	.087	.040
REPGPA	-.455	-.490	-.468	-.810	-.108

grouping (to mirror our usual state of knowledge in Context (D)), are contained in the lower triangular region. All values are reported to two digits only.

Table B.5 here

The fit is remarkable given the coarseness of the grouping methods in our example. Only one grouped estimate deviates by more than .02 from its corresponding ungrouped correlation coefficient. A statistical test of Goodness of fit would be superfluous even with a sample size of 2676.

The above example reflects the potential rather than the present. Firm guidelines for choosing the best grouped estimates for reconstructing a correlation matrix have not yet been developed. Here we knew the values needed and this enabled us to pick and choose among potential grouping methods. In practice the investigator with anonymously collected information on some primary variable is not so fortunate.

Table B.5 Individual-level correlation matrix reconstructed from standardized between-group regression coefficients.

	VARIABLES	SRAA	ACH	SAT	HSGPA	HSPHYS
		Individual-Level Coefficients (Upper Triangular Region)				
Best Estimates From Grouped Data (Lower Triangular Region)	SRAA	----	.53	.57	.37	.21
	ACH	.55	---	.84	.53	.32
	SAT	.58	.82	---	.49	.26
	HSGPA	.36	.52	.54	---	.13
	HSPHYS	.21	.34	.27	.13	---

APPENDIX B FOOTNOTES

[1]Since all variables are standardized, this is equivalent to examinig the zero-order correlation coefficients ρ_{YZ} and ρ_{XZ}. In fact, even when variables are unstandardized, it is better to compare correlations as it is the relative strength of relation that is important without regard to differences in variation of the variable.

[2]It is possible to specify the conditions under which $\beta_{YZ} > \beta_{YZ\cdot X}$ when X, Y, and Z are standardized.

$$\beta_{YZ} > \beta_{YZ\cdot X}$$

when

(i) $\beta_{YZ}\beta_{YZ\cdot X} < \beta_{YX}$ and β_{XZ} is positive, or when

or when

(ii) $\beta_{YZ}\beta_{XZ} > \beta_{YX}$ and β_{XZ} is negative.

In most cases, researchers will have some guidance as to whether these conditions hold even when data are collected anonymously.

REFERENCES

Affifi, A.A. and Elashoff, R.M., Missing Observations in Multivariate Statistics I: Review of the Literature. Journal of the American Statistical Association, 1966, 61, pp. 595-604.

Affifi, A.A. and Elashoff, R.M., Missing Observations in Multivariate Statistics II: Point Estimation in Simple Linear Regression. Journal of the American Statistical Association, 1967, 62, pp. 10-29.

Averch, H., Carroll, S.J., Donaldson, T.S., Kiesling, H.J. and Pincers, J., How Effective is Schooling?, The Rand Corporation, R-956-PCSF/RC, 1972.

Baird, L. and Feister, W.J., Grading Standards: The Relation of Changes in Average Student Ability to the Average Grades Awarded. American Educational Research Journal, 1972, 9, pp. 431-441.

Bidwell, C.E. and Kasarda, J.D., School District Organization and Student Achievement. American Sociological Review, 1975, 40, pp. 55-70.

Blalock, H.M., Aggregation and Measurement Error. Social Forces, 1972.

Blalock, H.M., Wells, C.S., and Carter, L.F., Statistical Estimation with Random Measurement Error. Pp. 75-103 in E. Borgotta and G. Bohrenstedt (Eds.) Sociological Methodology, San Francisco: Jossey-Bass, 1970.

Boruch, R.F., Assuring Confidentiality of Responses in Social Research. American Sociologist, 1971, 6, pp. 308-314.

Boruch, R.F., Maintaining Confidentiality of Data in Educational Research: A Systematic Analysis. American Psychologist, 1971, 26, pp. 413-430.

Boruch, R.F., Strategies for Eliciting and Merging Confidential Social Research Data. Policy Sciences, 1972.

Boruch, R.F., Social Research and the Confidentiality Issue: Methodological Perspectives. (Paper presented at the meetings of the American Educational Research Association, Chicago, 1972.)

Burkhead, J., Fox, T., and Holland, J.W., Input and Output in Large City High Schools. Syracuse: Syracuse University Press, 1967.

Burstein, L., Issues Concerning Inferences from Grouped Observations. (Paper presented at the Annual Meeting of the American Educational Research Association, April 19, 1974(a), Chicago, Illinois.)

Burstein, L., Consortium on Methodology for Aggregating Data in Educational Research. Proposal to the National Institute of Education, RFP-R-74-0009, 1974 (b).

Burstein, L., The Use of Data from Groups in Inferences About Individuals in Educational Research. Unpublished Ph.D. dissertation, Stanford University, 1975.

References, cont'd

Burstein, L., Kremer, D., and Gemoll, D., Identifying Effective and Ineffective Schools: A Case Study, in progress.

Coleman, J.S. et. al., Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Comber, L.C. and Keeves, Science Education in Nineteen Countries: An Empirical Study. International Studies in Evaluation, Volume I, International Studies in Evaluation, Sweden: Almqvist & Wiksell, Publishers, 1974.

Cramer, J.S., Efficient Grouping: Regression and Correlation in Engel Curve Analysis. Journal of the American Statistical Association, 1964, 59, pp. 233-250.

Duncan, D.D. and Davis, B., An Alternative to Ecological Correlation. American Sociological Review, 1953, 18, pp. 655-666.

Feige, E.L. and Watts, H.W., An Investigation of the Consequences of Partial Aggregation of Micro-Economic Data. Econometrica, 1972, 40, pp. 343-360.

Freeman, J. and Hannan, M.T., Growth and Recline Processes in Organizations. American Sociological Review, April, 1975, 40.

Glass, G.V. and Stanley, J.C., Statistical Methods in Education and Psychology. Englewood Cliffs: Prentice-Hall, 1970.

Goodman, L., Ecological Regression and the Behavior of Individuals. American Sociological Review, 1953, 18, pp. 663-664.

Goodman, L., Some Alternatives to Ecological Correlation. American Journal of Sociology, 1959, 64, pp. 610-625.

Hannan, M.T., Aggregation and Disaggregation in Sociology. Lexington, Mass., and Washington, D.C.: Heath, 1971.

Hannan, M.T., Data Aggregation in Educational Research: Recent Developments. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., April, 1975.

Hannan, M.T. and Burstein, L., Estimation from Grouped Observations. American Sociological Review, 1974, 39, pp. 374-392.

Hannan, M.T., Freeman, J., and Meyer, W., Comment on Bidwell and Kasarda. American Sociological Review, forthcoming.

Hanushek, E., The Education of Negroes and Whites. Ph.D. Dissertation, Massachusetts Institute of Technology, 1968.

Hanushek, E., The Value of Teachers in Teaching, RM-6362-CC/RC, The Rand Corporation, Santa Monica, California, December, 1970.

Hanushek, E.A., Education and Race. Lexington, Mass.: D.C. Heath, 1972.

Hanushek, E.A., Jackson, J., and Kain, J., Model Specification, Use of Aggregate Data, and the Ecological Correlation Fallacy. Political Methodology, 1974, 1, pp. 89-107.

Hauser, R.M., Schools and the Stratification Process. American Journal of Sociology, 1969, 74, pp. 587-611.

Hauser, R.M., Context and Consex: A Cautionary Tale. American Journal of Sociology, 1970, 75, pp. 645-654.

Hauser, R.M., Socioeconomic Background and Educational Performance. Rose Monograph Series, American Sociological Association, Washington, D.C., 1971.

Hauser, R.M., Contextual Analysis Revisited. Sociological Methods and Research, February, 1974.

Husen, T., International Project for the Evaluation of Educational Achievement. Phase I: International Study of Achievement in Mathematics: A Comparison of Twelve Countries. Almqvist and Wiksell, Stockholm, and John Wiley & Sons, New York, 1967.

Katzman, M., Distribution and Production in a Big City Elementary School System. Yale Economic Essays, 1968, 8, pp. 201-256.

Kiesling, H.J., The Relationship of School Inputs to Public School Performance in New York State. P-4211, The Rand Corporation, Santa Monica, California, October, 1969.

Kiesling, H.J., A Study of Cost and Quality of New York School Districts. Washington, D.C., Department of Health, Education and Welfare, 1974.

Klitgaard, R.E. and Hall, G., Are There Unusually Effective Schools? P-4995, Rand Corporation, April, 1973.

Levin, H.M., A New Model of School Effectiveness in Do Teachers Make a Difference?, OE-58042, U.S. Department of Health, Education and Welfare, 1970.

Lewy, A., Correlation Coefficients of Three Data Types: Zero Order Measures, Within Group Deviations, and Group Means. Unpublished Manuscript, Working Paper No. 16, University of Pittsburgh, 1972.

Madansky, A., The Fitting of Straight Lines When Both Variables are Subject to Error. American Statistical Association Journal, 1959, 54, pp. 173-205.

Mayeske, George W.; Okada, Tetsuo,; Cohen, Wallace, M.; Beaton, Albert, E.; Wisler, Carl E., A Study of the Achievement of Our Nation's Students. U.S. Government Printing Office, 1973.

Menzel, Herbert, Comment on Robinson's "Ecological Correlation and the Behavior of Individuals". American Sociological Review, 15:5 (1950), p. 674.

References, cont'd.

Michelson, Stephan, "The Association of Teacher Resourceness with Children's Characteristics", in How Do Teachers Make a Difference, U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Educational Personnel Development (OE-58042), 1970, 120-168.

Mood, A.M., Do Teachers Make A Difference?, U.S. Department of Health, Education and Welfare, U.S. Office of Education, Bureau of Educational Personnel Development, OE-58042, 1970, pp. 55-75.

Mosteller, F. and Moynihan, D.P. (Eds.) On Equality of Educational Opportunity. New York: Random House, 1972.

Peaker, G.F., An Empirical Study of Education in Twenty-One Countries. International Studies in Evaluation, Volume VIII. Sweden: Almqvist & Wiksell, Publishers, 1974.

Prais, S.J. and Aitchison, J., The Grouping of Observations in Regression Analysis. Review of the International Statistical Institute, 1954, 22, pp. 1-22.

Purves, A.C., Literature Education in Ten Countries: An Empirical Study. International Studies in Evaluation, Volume II. Sweden: Almqvist and Wiksell, Publishers, 1974.

Robinson, W.S., Ecological Correlations and the Behavior of Individuals. American Sociological Review, 1950, 15, pp. 351-357.

Rock, D.A., Centra, J.A., and Linn, R.L., Relationships Between College Characteristics and Student Achievement. American Educational Research Journal, 1970, 7, pp. 109-121.

Rock, D.A., Baird, L., and Linn, R., Interaction Between College Effects and Students' Aptitudes. American Educational Research Journal, 1972, 9, pp. 149-161.

Scheuch, E.K., Cross-National Comparisons Using Aggregate Data: Some Substantive and Methodological Problems, in R.L. Merritt and S. Rokkan (Eds.) Comparing Nations: The Quantitative Data in Cross-National Research, New Haven: Yale University Press, 1966, pp. 148-156.

Smith, M.S., Equality of Educational Opportunity: The Basic Findings Reconsidered, in F. Mosteller and D.P. Moynihan (Eds.) On Equality of Educational Opportunity. New York: Basic Books, 1972.

Summers, A.A., and Wolfe, B.L., Which School Resources Help Learning? Paper presented at the Annual Meeting of the Econometrics Society, New York, 1974.

Thorndike, R.L., Reading Comprehension Education in Fifteen Countries, International Studies in Evaluation, Volume III, Sweden: Almqvist & Wiksell, Publishers, 1974.

References, cont'd.

Walberg, H.J., Predicting Class Learning: An Approach to the Class as a Social System. American Educational Research Journal, 1969, 6, pp. 529-542.

Walberg, H.J. and Rasher, S.P., Public School Effectiveness and Equality: New Evidence and its Implications. Phi Delta Kappan, September, 1974, pp. 3-9.

Wiley, D.E. and Wiley, J.A., The Estimation of Measurement Error in Panel Data. American Sociological Review, 1970, 35, pp. 112-117.

ERRATA

Data Aggregation in Educational Research: Applications

Page 1--

Mayeske et al. should be dated 1973 . Also only one of the volumes is cited among the references.

The IEA studies are included in the references under the names of the authors (Husen, Comber and Keeves, Purves, Peaker, and Thorndike). Not all of the volumes are listed as some are still being revised and are not discussed in the paper. The correct publication dates are 1967 for the Husen volume, 1973 for the American edition of Comber and Keeves, Purves, and Thorndike, and hopefully, 1975 for Peaker.

Page 5--

The correct date for the Blalock reference is 1972 (Social Forces).
The correct date for the Wiley and Wiley is 1970 (American Sociological Review).

The first sentence of the last paragraph should read:
"Sound principles have already been developed..." ("been" left out of earlier draft.)

Page 6--

")" left off end of set of references in first 2 lines of page.

Page 8--

Delete "," after Scheuch's name in 2nd line from bottom.

Page 9--

The correct publication date for the Katzman reference is 1968.
")" after Levin (1970) in last paragraph.

Page 10--

The Summers and Wolfe (1974) reference for the Federal Reserve Bank of Philadelphia is in error. There are actually two reports from there so far. One appeared in Business Review (Summers and Wolfe) and the other was a presentation to the Econometric Society (presumably Summers). The comments in the paper on the study are based on discussions with Summers and a summary of the Econometric Society paper.

Page 12--

The last footnote to Table 2 ("f") reads " Ten year-olds were not included in the Literature study."

Page 14--

The heading "Individual Model" after "1." should read "Individual Level".

Page 16--

The Heading should read "CHOOSING THE APPROPRIATE UNIT OF ANALYSIS" (singular).

Lines 5 and 6 after heading-- sentence is missing "of", should read "...the specification of a set of research foci..."

Page 17--

Some versions are missing the reference to the Rand sponsored study by Klitgaard and Hall (1973) .

Page 19--

The study of administrative intensity cited on the 6th line is by Freeman and Hannan (1975) (Authorship position reversed).

Page 19 (cont.)

There has been some modification of the paragraph about the Walberg and Rasher (1974) study. Unfortunately , the modifications were made too late for inclusion in the duplicated version of the paper. The paragraph should read:

" For example, though their conclusions are appealing, Walberg and Rasher (1974) cannot avoid the methodological shortcomings of the Coleman study by using state-level data, especially with 1969 and 1970 selective service examination failures as the outcome variables. Walberg had previously cited Robinson's (1950) paper (Walberg, 1969, p.530) and thus should be aware of the danger of treating states as random groupings of persons. Besides, (a) women did not take military selective mental tests and (b) the years 1969-70 were not the times to try to pass the test (a case of the winners are the losers, in this writer's opinion). It is particularly unfortunate that this article appeared in a journal that is more widely read by administrators than by researchers since the former are less likely to realize its methodological limitations."

Page 20 --

Glass and Stanley's book has a publication date of 1970 rather than 1972.

Page 21--

The correct references to Hauser are 1969, 1970, 1971, 1974.

Page 22--

The subheading for the Lewy study should read "Group Anchored versus Sample Anchored Measures".

FOOTNOTES--

Footnote 2-- Citation should read See, e.g., Hauser (1974).

References--

The following citations have incorrect publication dates in earlier versions:

Comber and Keeves (1973)
Peaker (1975)
Purves (1973)
Thorndike (1973)

Also, there are unnecessary parentheses around the citations for Boruch (1972b) and Burstein (1974a).