DOCUMENT RESUME

ED 041 948 TM 000 043

AUTHOR Astin, Alexander W.
TITLE The Methodology of Research on College Impact.
SPONS AGENCY American Council on Education, Washington, D.C.; National Inst. of Mental Health (DHEW), Bethesda, Md.; National Science Foundation, Washington, D.C.
PUB DATE 70
NOTE 67p.; Portions of this paper were presented at the annual meeting of the American Educational Research Association, Minneapolis, Minn., March 1970

EDRS PRICE EDRS Price MF-$0.50 HC-$3.45
DESCRIPTORS College Environment, *College Students, Data Collection, Educational Change, *Environmental Influences, Institutional Research, Longitudinal Studies, Measurement Techniques, Reliability, Research Design, *Research Methodology, *Research Problems, Statistical Analysis, Student Development

ABSTRACT

Methodological difficulties inherent in the design of studies of college impact are reviewed and an attempt is made to provide an improved research design. The discussion of research problems focuses on those of inferring causation. A characteristic three-component model of student development, comprised of student output, student input, and college environment, is analyzed to illustrate the problems in the research design presently employed. Inferential errors resulting from the use of this design are categorized and discussed. Three statistical methods for analyzing the model components, matching, actuarial tables, and linear multiple regression analysis, possible measurement techniques, and various methods of data collection are individually evaluated. The paper concludes with suggested resolutions of the inferential dilemmas. (PR)

The Methodology of Research on College Impact[1]

Alexander W. Astin

American Council on Education

In their recent comprehensive review of research on college impact, Feldman and Newcomb (1969) summarize more than 1,000 empirical studies in which investigators have attempted to learn how students are affected by their college experience. For the most part, the findings from these studies are very difficult to interpret, primarily because of problems in research design and methodology. In view of the burgeoning state of current research on college impact, it may be useful to review these methodological difficulties and to suggest certain ways in which future research can be designed to avoid some of those which have plagued most of the studies reviewed by Feldman and Newcomb.

Among the problems that will be covered are the following: single-institution versus multi-institution studies, longitudinal versus cross-sectional data, alternative statistical designs, the effects of measurement error, alternative methods of measuring environmental variables, methods for detecting student-environment interaction effects, and methods of collecting data. Throughout the paper, however, the discussion will focus on problems of inferring causation: that is, of determining if and how the student is affected by his college experience.

A Conceptual Model

For purposes of discussion, we shall utilize a model of student development in higher education that has characterized much recent multi-institutional research.[2] In this model, the college can be seen as comprising three conceptually distinct components: student outputs, student inputs, and the college environment.

Student outputs refer to those aspects of the student's development that the college either does influence or attempts to influence. Although these outputs can be expressed at very high levels of abstraction (for example, "the utlimate welfare and happiness of the individual"), research is usually concerned with those relatively immediate outputs that can be operationalized. Specifically, then, the term outputs refers to measures of the student's achievements, knowledge, skills, values, attitudes, aspirations, interests, and daily activities. Adequate measures of relevant student outputs are, clearly, the sine qua non of meaningful research on college impact.

Remarkably, only a handful of the studies reviewed by Feldman and Newcomb were concerned with the impact of colleges on cognitive outcomes. There are, to be sure, many hundreds of studies of academic achievement (Feldman and Newcomb did not review these), but such studies are usually concerned with predicting college grade point averages rather than with measuring growth or change in cognitive skills or with assessing college impact on such skills. Considering that the development of the student's cognitive skills is probably the most common educational objective of both students and colleges, this lack of research is unfortunate. It can probably be explained by the logistical problems involved in measuring cognitive outcomes (the necessity for proctoring and the high costs of achievement testing, for example), problems much more formidable than those encountered in measuring attitudinal outcomes (which can be assessed with relatively inexpensive, self-administered questionnaires).

Student inputs are the talents, skills, aspirations, and other potentials for growth and learning that the new student brings with him to college. These inputs are, in a sense, the raw materials with which the institution has to deal. Many inputs can be viewed simply as "pre-tests" on certain outputs (career choice and personal values, for example), whereas others (sex and race, for example) are static personal attributes. Inputs can affect outputs either directly or by interaction with environmental variables.

The college environment refers to those aspects of the higher educational institution that are capable of affecting the student. Broadly speaking, they include administrative policies and practices, curriculum, physical plant and facilities, teaching practices, peer associations, and other characteristics of the college environment.

The relationships among these three components of the model are shown schematically in Figure 1. The principal concern of research on

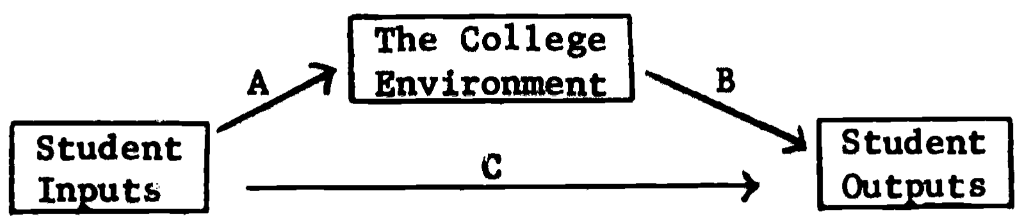


Figure 1

college impact is to assess relationship "B," the effects of the college environment on relevant student outputs. Relationship "C" refers to the fact that outputs are also affected by inputs, and relationship "A" to the fact that college environments are affected by the kinds of students who enroll.

In addition to the "main" effects of college environments on student outputs (B), the investigator may also be interested in certain interaction effects involving student inputs and college environments.

The diagram suggests that there are two types of interaction effects: those in which the effect of input on output is different in different college environments (AC), and those in which the effect of the college environment is different for different types of students (AB). Research on college impact is ordinarily concerned more with the second type.

Problems of Design

Although the ideal study of college impact would incorporate information on all three components of the model -- student inputs, environments, and student outputs -- most of the studies covered by Feldman and Newcomb lacked data on at least one of these components. In this section, we shall review some of the inferential problems that characterize such studies.

Following the conventions of statistical inference, we can assume that studies of college impact should be designed to minimize two kinds of inferential error:

Type I errors (rejection of the null hypothesis when it is true) occur when there is no college effect, but the investigator concludes that there is.

Type II errors (acceptance of the null hypothesis when it is false) occur when there is a significant college effect, but the investigator concludes that there is not.

The special problems inherent in the design of college effects studies indicate that there is still a third type of inferential error which we shall call Type III errors. These occur when there is a significant college effect, but the investigator concludes that the opposite effect occurs.[3] In a sense, a Type III error combines both Type I and

Type II errors, since it involves simultaneously the rejection of a null hypothesis which is true and (implicitly) the acceptance of a null hypothesis which is false. (A convenient mnemonic device for defining Type III errors is that 1 + 2 = 3.)

Some of the controversy over the design of college impact studies stems not so much from basic disagreements over design strategy as from differences in the relative values assigned to Type I and Type II errors. Investigators who are primarily concerned about minimizing Type I errors, for example, fear that the highly nonrandom distribution of students among institutions will lead educators and students to conclude that certain college "effects" exist when, in fact, they do not. Thus, they regard adequate control of differential student inputs as an essential feature in their design. Researchers who are more concerned about Type II errors, on the other hand, fear that too much control over student inputs will reduce the chances of finding environmental effects. These two somewhat opposed emphases are in part historical. That is, the earliest investigators of college impact exerted virtually no control over student inputs; as a consequence, the very substantial institutional differences in student outputs which were found they attributed to the environmental influences of the colleges. When another group of investigators subsequently re-examined this early work, they discovered that differences in institutional outputs could be largely attributed to differences in inputs and that the relative "impacts" of colleges diminished markedly once these differential student inputs were taken into account. Most recently, however, some investigators have been disturbed by the possibility that a design which controls for student inputs may tend to underestimate the differential impact of colleges or, in some cases, to obscure particular environmental effects.

We shall discuss these and other possible effects of various designs on both types of error in the section on Multi-institution Longitudinal Studies (below). It should be pointed out here, however, that very large differences in student inputs at various institutions -- a prominent characteristic of higher education in the United States (Astin, 1965b) -- are almost sure to make for large differences in student outputs, regardless of the actual effects of institutions. As a result, failure to take into account these differences in input when studying college effects virtually guarantees that the investigator will commit some Type I errors. More important, ignoring differential student inputs maximizes the investigator's chances of committing Type III errors.

Studies of "Growth" or "Change" at Individual Colleges

Perhaps the prototypical study reviewed by Feldman and Newcomb involves the testing and retesting of students at a single institution. Characteristically, the students complete an attitudinal questionnaire or inventory when they first enter college and take it again one year later, four years later, or in a few cases, many years after graduation. Measures of "change" or "growth" are obtained by comparing the student's input scores from the initial administration with his output scores from the followup administration. (These comparative measures are usually simple difference scores, although residual gain scores are used occasionally.) In subsequently interpreting these scores, the investigator typically assumes that any observed changes are due to the students' experiences in college. In other words, he equates "change" with "impact."

This type of design has the advantage of focusing attention on the longitudinal nature of student change and development in that it views the student's output performance in relation to his input characteristics. Its glaring weakness, however, is that it really produces no information that bears directly on the question of environmental impact. Would the same changes have occurred if the student had attended a different kind of college or had not gone to college at all? In the context of our conceptual model, this type of study yields information on student inputs and outputs but not on the environment. Thus, the college environment is not a variable but a constant. (The situation here is identical to the one encountered in experimentation when no control group is used.)

The very practical danger in assuming that change equals impact can be illustrated with an anecdote. I recently overhead a colleague from a highly selective small college complaining that nearly a third of his undergraduates who start out majoring in science shift to a nonscience field before graduation. He interpreted this decline in science interest (change) as somehow resulting from the science curriculum of the college (impact). As a consequence, he and other members of a committee on curriculum reform were seriously considering major changes in the science curriculum of the college in the hope of reducing the number of students who withdraw from science fields. As it happened, this colleague's institution was one of several hundred colleges participating in a longitudinal study of institutional impact on career choice (Astin & Panos, 1969). What he did not know was that the longitudinal analyses had revealed that the dropout rate from science was actually lower at his college than at almost any other college in the sample. Thus, his college was exerting a relatively positive rather than a negative influence on the student's interest in science. Under these circumstances, major changes in the existing science curriculum could very well increase rather than decrease the student dropout rate from science at the college.

Many investigators use a variation of this basic design: instead of collecting longitudinal information, they simply compare groups of freshmen and upperclassmen simultaneously on some measure. This method is so full of pitfalls, (many of which are discussed at length by Feldman and Newcomb) that one wonders if there is the slightest justification for supposing that the observed "changes" are in any way related to the college experience. In addition to the problems already mentioned, this method carries with it potentially serious deficiencies in sampling. It rests on the assumptions that (a) upperclassmen are a representative sample -- at least insofar as the output variable is concerned -- of the total cohort of freshmen from which they were drawn, and (b) this original cohort was drawn from the same population as the current freshmen who are being compared with the upperclassmen.

The tenuousness of these assumptions is obvious when one realizes that any sample of upperclassmen necessarily excludes dropouts and includes transfers -- two groups that are very likely to differ from the students who entered as freshmen and continued on without a break in their undergraduate progress. Moreover, changes in the nature of successive entering freshman classes may occur as a result either of modifications in either the applicant pool or admissions practices or of changes in the college student population itself. That such population shifts are indeed possible -- even over a brief period of time -- is revealed by the ACE's annual surveys of entering freshmen.[4] For example, during the four most recent years -- 1966, 1967, 1968, and 1969 -- the percentages of entering freshmen who checked "none" as their present religious preference has gone up consistently: 6.9, 7.9, 9.6, and 13.2 These trends held true for both men and women and for students at most types of institutions. Thus, even if no students had dropped out or changed their religious preferences since entering college, a comparison of the current (1969-1970) freshmen with the current senior classes at many colleges would lead to the conclusion that nearly half of the students who initially reported that they had no religious preference "changed" to some other choice after entering college.

The dangers in assuming that "change" is equivalent to "college impact" suggest that changes in students during college should be viewed as comprising two components: change resulting from the impact of the college and change resulting from other influences (maturation, non-college environmental effects, etc.). Note that the college may (a) bring about changes which otherwise would not occur, (b) exaggerate or accelerate changes resulting from other sources, or (c) impede or counteract changes resulting from other sources (as in the example, cited above, where the college's dropout rate from science was much lower than average).

Studies of Environments and Student Outputs

One alternative to the single-institution studies that are so common in research on college impact is the multi-institution study, in which the student outputs of several institutions are compared. It was a frequent practice during the 1950's, for example, to compare institutions on such output measures as the percentage of graduates obtaining Ph.D.

degrees or the number of alumni listed in Who's Who. While such studies have the advantage of permitting the investigator to study variations in college environments, the empirical findings that result tend to be highly ambiguous because the student input has been disregarded.

The importance of student input data in multi-institution studies is aptly illustrated by the history of research on "Ph.D. productivity." The earliest of these studies indicated that the graduates of certain colleges and universities were much more likely than were the graduates of other institutions to win fellowships for graduate study and to go on to obtain the Ph.D. degree (Knapp & Goodrich, 1952; Knapp & Greenbaum, 1953). More important, the environments of the "highly productive" institutions, when compared with those of the less productive ones, were found to have higher faculty-student ratios, larger libraries, more funds for scholarships and research, and similar resources usually assumed to indicate institutional "excellence" and eminence. In short, the causal inferences drawn from these early studies were that such institutional resources are conducive to the development of the student's motivation to seek advanced training. Among other things, this research evidence seemed to confirm the folklore about what makes for "quality" in higher education. Taken at face value, and assuming that the output measure under study (motivation to seek advanced training) was relevant to the goals of the institution, these findings offered empirical support to the administrator in his attempts to increase the size of his faculty, library, and so forth.

But the validity of these earlier studies came to be doubted when it was shown that institutions differ widely in their student inputs: Highly productive institutions, for example, enroll greater proportions of academically able students than do less productive institutions (Holland, 1957). Intellectually advantaged students are, of course, more likely than are average students both to win graduate fellowships and to be interested in pursuing the doctorate <u>even if their institution exerts no special influence during the undergraduate years</u>. These doubts were subsequently confirmed by a series of studies (Astin, 1962, 1963a, 1963b) in which differential undergraduate student inputs to diverse institutions were controlled. Thus, when the abilities, career plans, and socioeconomic backgrounds of the entering students were taken into account, an institution's output of Ph.D.'s was revealed to be largely a function of the characteristics of its entering students rather than of its resources. Moreover, certain types of institutions that were earlier described as "highly productive" of Ph.D.'s turned out to be underproductive in relation to their student inputs. In addition, the apparent "effects" of library size, faculty-student ratio, and other similar indicators of institutional quality disappeared.

Multi-institution Longitudinal Studies

The inferential problems inherent in single-institution studies and in those multi-institution studies that do not utilize student input data indicate that an adequately designed study of college impact requires information concerning all three components of our model: student inputs, college environments, and student outputs. Merely collecting such data, however, does not assure that true college effects will be identified and spurious college effects will not. The avoidance of such inferential errors depends on a number of factors, including the nature of the student input data obtained and the statistical method used to analyze the data.

Since there is no way to guarantee that the nonrandom distribution of students among institutions will be compensated for completely, *the investigator's task in collecting data and in selecting a statistical method is simply to reduce the chances that his inferences will be wrong*.

Student Input Data

"Relevant" student input data are those which affect the student's choice of a college or the student output variable under study or both. To reduce the chances of committing Type I errors, however, it is not necessary to collect both types of data: As Figure 1 indicates, an unbiased estimate of the environment-output effect (B) can be obtained if either A (input-environment) or C (input-output) is controlled. If both relationships are controlled, however, a more sensitive test of environment-output effects will result, thus reducing the probability of committing Type II errors.

The designs that result when student input data are controlled in different ways can be depicted by a simple 2 x 2 table (Figure 2). (For purposes of illustration, we have used the terminology of linear multiple regression to label the different designs, although the basic logic of the designs does not require that linear regression be the method used.)

		Input Partialled Out of Output?	
		Yes	No
Input Partialled Out of Environment?	Yes	Partial Correlation	Part Correlation I
	No	Part Correlation II	Zero-Order Correlation Between Environment and Output

Figure 2. Four Types of Multi-institution Designs for Studying the Relationship Between College Environments and Student Outputs.

The lower right-hand box in Figure 2 repre ents the multi-institutional design in which no student input data are used (see previous section for examples). The upper left-hand box represents the partial correlational design in which the effects of student inputs on both output and environment are controlled. As we have already indicated, this design provides the most sensitive test of environmental effects.

Part correlation II (lower left box of Figure 2) involves control over the input-output relationship but not over the input-environment relationship.[5] Since the total output variance is likely to be more dependent on input than on environment, this design is probably the second most sensitive of the four. An interesting application of this design in multi-institution studies is first to solve the regression equations using all students, and then to aggregate the residual output scores of students within an institution, thereby producing a mean residual output score for the institution. This mean residual can provide a useful quantitative measure of institutional "impact." For example, in a recent multi-institution longitudinal study (Astin & Panos, 1969), one of the output measures was whether or not the student had dropped out during the four years after entering college (scored as a dichotomy: 1 = stayed in college for four years; 0 = dropped out). The mean college residuals on this measure, which varied among the 246 colleges from -30% to +16%, thus provided a measure of the extent to which each college's retention rate was either above or below what would have been expected from the characteristics of its entering students.

The fourth design shown in Figure 2 -- part correlation I -- involves control over the input-environment relationship but not over the input-output relationship. Although this method has seldom been used, it yields an estimate of the efficiency of a particular environmental variable (or combination of environmental variables). By "efficiency" we mean the extent to which the total variation in student performance on the output measure can be attributed solely to theoperation of environmental variables.

A final question relating to student input data concerns the data which are actually used. Investigators who are used to thinking in terms of experimental rather than correlational models run the risk of utilizing only a single student input measure, a "pretest" measure or "covariate." (Investigators who regard research on college impact as simply a matter of "change" will be similarly tempted to rely on a single input measure.) The problem with single input measures is that they are almost sure to be inadequate, since the distribution of student inputs among institutions is biased with respect not just to one but to many student attributes (Astin, 1965b; Holland, 1959). Because the factors influencing college choice are often difficult to identify, probably the best protection for the researcher here is to measure and control all student attributes that are likely to affect the output measures under study.

That using only a single "pretest" measure in studying college impact can seriously bias the conclusions is illustrated by a recent study in which the three Area Tests on the Graduate Record Examination (GRE) were used as output measures (Astin, 1968b). In all three analyses, the student's initial ("pretest") aptitude entered as the first variable in the stepwise regression. In two of the three analyses, however, the student's sex entered with a large weight at the second step. The sex ratio in the student body, it should be noted, is also strongly related to environmental attributes such as Cooperativeness, Cohesiveness, and Femininity (Astin, 1968a). Clearly, if initial aptitude had been the only student input variable considered, the findings might have shown -- incorrectly -- that the student's learning and achievement are significantly affected by the degree of cooperativeness, cohesiveness, or femininity of the college environment.

Statistical Alternatives

Three basically different statistical methods have been used to analyze input, output, and environmental data: matching, actuarial tables, and linear multiple regression analysis.

Matching. Perhaps the least desirable statistical approach is to match students entering different colleges in terms of their input characteristics. Not only are many subjects lost, but also the subsamples of students selected for study are unrepresentative of their institutions; thus potentially serious regression artifacts are introduced into the data. These and other problems with matching designs have been discussed at length elsewhere (e.g., Campbell & Stanley, 1963). Briefly, the major inferential problem is that the analysis is likely to yield artifactual "effects" which are in reality the result of errors of measurement in the input variables. (These and other problems associated with measurement errors will be discussed in the next section.)

It is not generally recognized that research on interaction effects ordinarily employs a kind of matching design and is therefore liable to the same deficiencies. That is, if one sorts out his students in terms of attributes such as ability, sex, race, and so forth, his findings on the effects of college variables are likely to be biased by errors of measurement in these student attributes. For example, if students are stratified by ability level (in order to detect possible interaction effects between ability and some college characteristic), the students in any given ability category will include representative subsamples of students from some colleges and highly nonrepresentative subsamples from others. Since the error of measurement in any student's ability test performance is likely to be correlated with the extent to which is score deviates from the mean score of his classmates (i.e., students with relatively high scores being more likely to have positive errors of measurement than students with relatively low scores), the "superior" students from the least selective colleges are more likely to have spuriously high ability test scores (i.e., positive measurement errors) than are the "superior" students from the most selective colleges. Thus, such studies may find that college selectivity, or some variable correlated with selectivity, "affects" highly able students, when in fact there is no effect. (For a fuller description of this phenomenon, see the next section.)

Even interaction studies that use highly objective student input characteristics, such as race and sex, are not free from these artifacts, since the measurement of these attributes too is likely to contain some error: A few women will probably be misclassified as men, a few nonwhites misclassified as whites, and so forth. The <u>bias</u> occurs because such errors are not equally probable in all types of institutions. Thus, there is likely to be more error of measurement in classifying students as "white" (i.e., more nonwhites) among those attending predominantly Negro colleges than among those attending predominantly white colleges. Similarly, the chances are probably greater that students classified as "female" are really male if they are attending a technological institution than if they are attending a teachers college.

Actuarial Tables. The second basic type of design involves the use of actuarial tables for controlling differential student inputs. Actuarial tables are especially helpful when the input variables are qualitative rather than quantitative in nature (Astin, 1962, 1963a). Briefly, what the investigator does is to sort his total pool of subjects into discrete cells on the basis of their input attributes (by sex, by race, by family SES, and so forth). Cells need not, of course, be balanced (e.g., one might form separate sex cells for one race but not for another). The purpose of the sorting procedure is to generate new cells in such a way that the between-cell variance in the output measure is maximized and within-cell variance is minimized. The actuarial approach is similar in some ways to multiple group discriminant analysis, except that the roles of the independent and dependent variables are reversed. In discriminant analysis, the groups define the dependent variable and the independent variables are used to form a metric which maximally discriminates the groups. Conversely, in actuarial analysis, the metric is a given (the dependent variable), and the independent variables are used to form groups which maximally discriminate this metric.[6]

At the point where it is no longer possible to form further groups which significantly discriminate with respect to the output variable, the "predicted" or "expected" output score for each student becomes the mean output score of all students occupying the cell where he is located. (One refinement, which might be desirable for use with relatively small samples, is to exclude a given subject in computing the mean for his cell.) The difference between the expected value and the student's actual score on the output measure thus becomes the dependent variable for analysis of college impact. Students can be sorted into their respective colleges, and the mean discrepancy score computed separately for each college. At this stage, one has a situation similar to that described above for part correlation II (Figure 2).

The principal advantages of the actuarial table are that it is probably much easier to understand than standard regression techniques (below) and that it permits the investigator to take into account interaction effects among student input variables as well as nonlinear effects of input variables. Its principal disadvantage is that there is no generally accepted analytic method for determining how the cells should be formed; the number of possible cells increases exponentially as the number of input variables increases. Moreover, to use the method effectively, one must have a very large sample of subjects. Nevertheless, it has been shown (Astin, 1962, 1963a) that actuarial tables can produce wide separations on student output measures and that, given large enough samples, the cell means prove to be highly stable on cross-validation. While the actuarial approach does not easily accomodate variables that are continuous (rather than qualitative), it is possible, under certain conditions, to combine actuarial tables with regression analyses (see below).

Linear Multiple Regression. The statistical method used most frequently in recent multi-institution longitudinal studies is linear multiple regression. This technique can be applied in three basically different ways. The first, ard perhaps the most straightforward, approach is the "full model" described by Bottenberg and Ward (1963), in which the output measure is regressed on both input and environmental variables, with the student used as the unit of analysis.

The second application is identical to the first, except that the institution rather than the student is used as the unit in the analyses both of environmental effects and of input effects. Mean scores on each input variable are calculated separately for each institution. Then the output variable is regressed on these mean input variables, after which the environmental variables are permitted to enter the analysis. Although the much smaller number of units involved (institutions versus students) makes this method computationally much simpler than the first, it should be used with caution because it greatly increases the probability of type II errors. The major problem here is that the method treats peer group effects as input, rather than environmental, effects. That is, many potentially important environmental variables are a reflection of (or at least highly correlated with) the aggregate or mean score on particular student input characteristics (the mean ability level of the student body, for example). If the magnitude of a particular environmental effect is proportional to the institution's mean score on a particular input variable, this method may partial out the environmental effect along with any input variable effects.[7]

The third application of regression analysis, which was alluded to in the previous discussion of actuarial tables, combines the first two uses. The output variable is first regressed on input variables using the student as the unit of analysis. Mean residuals are then computed separately by college, and the effects of environmental variables assessed using the institution as the unit of analysis. This application is useful when the magnitude of the effects of particular colleges is being investigated. Moreover, in cases where the output measurement has a meaningful zero point (the percent of students who are affected in some way, for example), the third method offers some interesting possibilities for further analyses of the mean residuals. For example, one might conduct a series of analyses using several different outcome criteria, each of which has been scored dichotomously so that the mean residual indicates the percentage of students at a given college who were differentially affected by that college. An empirical typology of colleges could then be developed by factoring the covariances (rather than the correlations) among the mean residuals obtained on the several different output variables. The resulting typology would thus give greatest weight to those environmental influences that affect the largest percentage of students (i.e., to those output variables whose mean residuals show the greatest interinstitutional variance).

A similar two-stage analysis can be carried out using the first two regression methods. Student input variables are first permitted to enter the regression equation, after which environmental measures are permitted to enter. Some investigators object that such a two-stage analysis biases the findings in favor of student input, as opposed to environmental, variables, but this supposed "rivalry" between the two types of variables is something of a straw man. Student input variables are controlled prior to the assessment of environmental effects for two reasons. First, there is the practical problem of reducing type I and type III errors. Unless some control over differential student inputs is exerted prior to the assessment of environmental effects, the investigator maximizes his chances of committing both of these types of inferential errors. Second, there is the logical question of the temporal sequence of student input and environmental variables. While the college environment clearly can be influenced by the nature of the student input, it is illogical to assume that the student's input characteristics have been affected by the environment of his college. That is, the student's sex, race, SES, initial aptitude, and other input variables are set before he has any opportunity to be exposed to the college environment. It is true, of course, that the entering student's plans or attitudes may already have been influenced by his expectations about the college or by his having been accepted by the college for admission, but we are concerned here with the "environmental effects" that occur only after he matriculates.

One way of regarding the problem of what percentage of the total output variance can be attributed to input or environmental variables is to conceive of the predictable variance in any output measure as comprising three conceptually and statistically separate components:

$R^2_{PART\ (I \cdot E)}$. <u>The percentage of output variance uniquely attributable to student input variables</u>. This quantity refers to Part Correlation II (Figure 2) and is the squared multiple correlation between the output measure and the residual input variables (i.e., input independent of environmental variables).

$R^2_{PART\ (E \cdot I)}$. <u>The percentage of output variance uniquely attributable</u>

to college environmental variables. This quantity refers to Part Correlation I (Figure 2) and is the squared multiple correlation between the output measure and the residual environmental variables (i.e., environmental independent of input variables).

$R^2 - R^2_{PART\ (I \cdot E)} - R^2_{PART\ (E \cdot I)}$. The percentage of confounded output variance, or output variance which is jointly attributable to input and environment.

The first two coefficients provide "lower-bounds" estimates of the total output variance that can be attributed, respectively, to input and environmental sources. An alternate method of computing "lower-bounds" estimates would be simply to determine how much R^2 increases when one set of variables is added to the other. This latter approach, however, may give too high an estimate because it would assign all "suppressor" effects[8] between the two sets of variables to the second set (i.e., the set being added to the equation).

"Upper-bounds" estimates can be obtained simply by adding the lower bounds estimates to the confounded variance. An alternative method would be simply to leave out one set in computing R^2. This alternate "upper-bounds" estimate, however, may be too low because it does not capitalize on any possible suppressor effects between environmental and input variables.

During the past several years, some discussion has appeared in the literature concerning the most appropriate use of multivariate analysis in analyzing student input, environmental, and student output information. Some writers (Werts & Watley, 1969) prefer to pool all input and environmental variables in a single analysis rather than to use the two-stage input-environment analysis. The resulting regression coefficients, according to these writers, would reflect the "independent contribution" of various input and environmental variables in accounting for variation in the output variable. One interpretive difficulty with this method is that the various input and environmental variables are not independent. Under such conditions, some writers have concluded that "the notion of 'independent contribution to variance' has no meaning when predictor variables are intercorrelated (Darlington, 1968, p. 169)." The problem here is essentially one of what happens to the confounded variance. Since this variance must be reflected in the regression coefficients, there is no way to determine erely from these coefficients just how much of the confounded versus unique variance has been allotted to any independent variable or class of variables. Another problem is that the regression coefficients do not show whether a particular variable is acting directly on the output variable or whether it is operating primarily as a suppressor variable by accounting for extraneous variance in other independent variables.

A possible solution to these problems associated with regression weights would be to compute "lower-bounds" estimates of the unique influence of a particular variable or class of variables by means of squared part correlations and then to compute "upper-bounds" estimates by adding the confounded variance to the squared part correlations.[9] The investigator could then evaluate these two estimates in terms of the various risks that he is willing to take of incurring type I and type II errors. Obviously, the greater the discrepancy between the upper- and lower-bounds estimates, the greater the risks.

A variation in the use of regression analysis proposed by some authors is causal path analysis (Duncan, 1966). Perhaps its major advantage over ordinary linear regression analysis is that it forces the investigator to specify the known or hypothesized relationships among his input, environmental, and output variables and aids him in differentiating "direct" from "indirect" influences on output variables. Path diagramming can also be a useful way of helping the investigator to see possible connections among his variables that he had not considered previously (Werts, 1968). Perhaps the major limitation of this method is that it can be unwieldly or even unworkable when the number of independent variables is large or when their temporal sequencing is not known. Since path analysis is more useful for testing specific causal hypotheses than for an open-ended exploration of college impact, its use should probably be confined to situations where the number of independent variables is relatively small, and where their interrelationships are relatively well understood.

Causal analyses of input, output, and environmental data by means of multiple regression techniques as a general approach to studying college impact has been criticised by Richards (1966) (who is also cited at length by Feldman and Newcomb, Appendix F) on the grounds that residual values are "notoriously unreliable and subject to errors of various sorts." But residual values are no less "reliable" than difference scores or even the change scores which Richards himself recommends as alternatives. Richards also objects to the regression approach because it can "obscure true college effects." In this regard, it is important to note that the first and third uses of regression analysis described earlier (i.e., the ones where inputs are controlled using the student as the unit of analysis) will not "obscure" even very small environmental effects, except in the special case where there is total confounding of input and environmental variables. (It would be difficult, for example, to compare the effects of men's colleges and women's colleges on some output variable that is related to sex.) Under these circumstances, there is no within-college variance in input, so that the environmental and input variables are completely confounded. However, as long as there is some overlap between institutions in student input characteristics, the application of regression analysis described above will not obscure any college effects, no matter how small.

Detecting Interaction Effects

In presenting our three-component model of college impact, we indicated that at least two kinds of student-college interaction effects can occur: those in which the effect of input on output is different in different college environments and those in which the effect of the environment is different for different types of students.[10]

In certain respects, the problem of interaction effects between student and college characteristics has more practical significance for administrative policy than the problem of the main effects of college environmental variables. A knowledge of environmental main effects is useful only when it is possible to modify existing colleges or to design new colleges in ways which will maximize the desired main effects. A knowledge of interaction effects, however, can be useful if there is no realistic possibility of making significant changes in existing college environments, since such knowledge permits one to maximize desired educational objectives by redistributing students among existing institutions in the most efficacious way. Such knowledge is of obvious value to large city or state systems comprising several institutions; it can also be useful to individual private colleges in selecting students who are likely to benefit most from the particular program offered by the institution.

Knowledge concerning interaction effects can also be applied within individual institutions. It can be used, for instance, as a basis for selecting those students most likely to profit from counseling and guidance in situations where resources for these services are limited, or for assigning students to various schools and colleges within an institution. Even if the final decision is left to the individual student, information about interaction effects can help him to make the most appropriate choice.

Assessing interaction effects presents many methodological problems, primarily because the number of possible student-environmental interaction effects is so large. Simply to "shotgun" the study of interaction effects by generating all possible combinations is usually unrealistic, either because of the large loss in degrees of freedom or because of limits on the number of variables that can be accommodated in a given analysis.

Perhaps the most common approach is to generate only those interaction terms suggested by a particular theory. However, the paucity of comprehensive theory in this field greatly limits the range of interaction terms that one can explore in this manner.

Another approach is to select a limited number of student input variables on the basis of their intrinsic importance (sex, race, SES, ability, for example) and to determine which are most likely to interact with environmental variables across a wide range of student outcomes. Future studies of college impact could then routinely examine interaction effects involving such variables.

There are many possible methods for assessing interaction effects in the multivariate model. The simplest (and probably most expensive) is to perform separate analyses on subgroups of students (all men, for example) defined in terms of the student characteristics that might interact with environmental variables. Another approach is to perform only one analysis, but to "score" the interaction terms (student ability x college size, for example) as a separate variable. This method has the advantage of computational simplicity, provided that the number of interaction terms is not excessively large. If the investigator wishes to assess such interaction terms using a very large number of environmental variables, the former method of separate analyses by subgroups is probably preferable.

Whatever method the investigator uses, he cannot be sure he has identified significant interaction effects until he has first controlled for the main effects of the variables that make up the interaction term. (The problem here is similar to the one encountered in analysis of variance designs, where the main effects of the independent variables must first be removed before the interaction effects can be studied.)

Many investigators who study college impact fail to recognize the need for controlling main effects before examining interaction effects. Take, for example, studies of the "congruence" between the student and his college, which are designed to test the assumption that the student's success and his satisfaction with college will be related to the degree of similarity between certain of his own characteristics and some comparable measure of the college environment. In some cases, the student's personality is compared with the typical personality of his fellow students; in others, his expectations about the college environment are compared with some "objective" measure of that environment. Whatever the measure used, such studies do not yield evidence on the importance of congruence unless one first examines how the student input and the environmental variables directly affect the output under consideration. In the single-institution study, of course, the main effects of the environmental variables cannot be tested since the environment is essentially a constant rather than a variable. Even in multi-institution studies, however, the main effects of the student input and the environmental characteristics in question have first to be controlled.

The stepwise linear regression model provides a convenient way of examining such interaction effects. This method allows the investigator to score his interaction terms as separate variates, omitting them from the stepwise analysis until the significant main effects of the input and environmental variables are controlled.

A similar multistage regression analysis can be used to assess other kinds of interaction effects. For example, if the investigator wishes to analyze possible interaction effects <u>among</u> input variables or <u>among</u> environmental variables as well as those between input and environmental variables, his analysis would involve separate stages in which the various effects would be controlled in the folowing sequence: main effects of input variables, interactions among input variables, main effects of environmental variables, interaction effects among environmental variables, and interactions between input and environmental variables.

The assessment of interaction effects by repeated analyses of different subgroups rather than by the multistage analysis just described presents certain problems for the investigator in determining when he has actually identified a significant "interaction" effect. Some investigators are tempted to conclude that they have done so in cases where a particular environmental variable is found to have a significant effect in one subgroup but not in another. It would be more definitive to test the significance of difference between environment-output correlations rather than simply to ascertain that one correlation is significant but not the other. Even under these conditions, however, difference in the sizes of the samples or in the variances of either environmental or output variables can affect differences between environment-output correlations. The method of separate analyses by subgroups also presents the problems discussed in the previous section in connection with matching designs.

Error of Measurement

We have already indicated that one major difficulty with matching designs is the bias caused by error of measurement in the matching variables. What happens is that, in order to match subjects on an attribute like, say, academic ability, unrepresentative subsamples of students must be selected from each institution. If the institutions under study differ markedly in selectivity, the investigator must select the least able students from the highly selective institutions and the most able students from the least selective. Since the latter are above average in ability in comparison with their classmates, their scores are likely to contain more positive than negative errors of measurement. By contrast, the students from the highly selective colleges, being below average in ability relative to their classmates, are likely to have more negative than positive errors of measurement in their scores. Under these conditions, in each matched pair of students, the one from the highly selective college will more often than not have a higher "true" score than the one from the least selective college. Another way of demonstrating this effect is to give all members of each matched pair a second, independent test. Students from the highly selective colleges will tend to score slightly higher than they did on the first test, whereas students from the least selective colleges will tend to score slightly lower than previously. Multivariate analyses avoid such errors by making it possible to utilize all subjects from all institutions, so that within-institution errors of measurement in the input variables sum to zero.

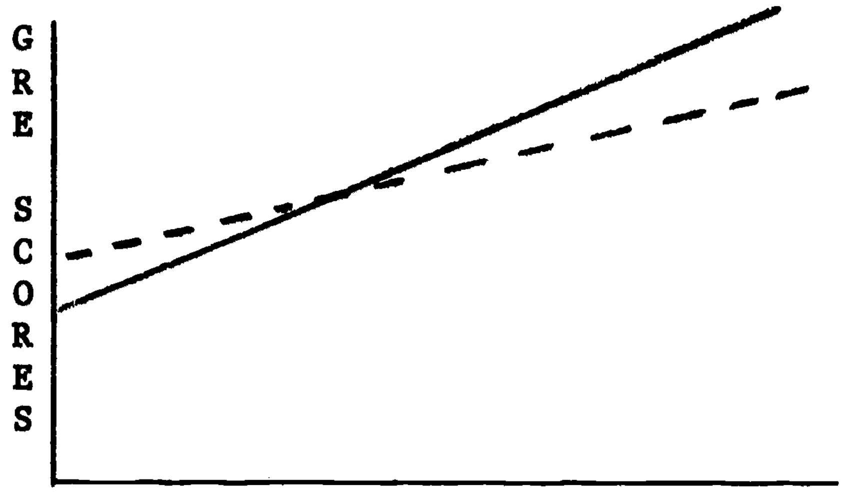
Measurement error can introduce the opposite kind of bias if the fallible input measure is used also as a basis for selecting students for admission. Under these circumstances, students in the highly selective institutions will tend to have more positive than negative errors of measurement in their test scores, whereas the rejects from these colleges, who enter less selective institutions, will tend to have more negative than positive errors in their scores. Unless the investigator can estimate precisely the extent to which a particular measure is relied upon in the admissions process, he would do well to avoid using the measure altogether in his longitudinal analyses of college impact.

An even subtler and potentially more serious consequence of measurement error -- one that arises not only in matching designs but also in designs that use correlational methods -- is attenuation in the observed correlation between input and output variables. It is a well-known statistical fact that error of measurement in either of a pair of correlated variables lowers their observed correlation. To the extent that such attenuation results from error in the input variables, a serious bias is introduced into the analysis if the input variable is also correlated with environmental variables.

The way in which this bias operates can be indicated by means of a hypothetical example. Assume that we are interested in determining how the student's achievement is affected by the "quality" of his college. Furthermore, we have longitudinal data on students attending a variety of colleges of differing degrees of quality. Our output measure of achievement is the student's composite performance on the Graduate Record Examination (GRE), and our environmental measure of quality is the percentage of faculty members holding Ph.D.'s. For the sake of simplicity, let us assume that there is only one relevant input measure: the student's composite score on the National Merit Scholarship Qualifying Test (NMSQT). To complete our hypothetical picture, we can make the following additional assumptions: (a) The NMSQT was not used in making admissions decisions; (b) The NMSQT is positively correlated with the GRE; and (c) NMSQT scores are positively correlated with college quality. This last assumption states simply that bright students are more likely to attend relatively high-quality colleges.

Our analyses of these three measures might involve simply computing the partial correlation between college quality and GRE performance, holding constant the effects of NMSQT scores. Or perhaps we might want to compute residual GRE scores (regressed on NMSQT scores) and plot the mean residuals for each college against the college quality measure (this latter type of analysis would permit us to see any non-linear effects of quality and also to identify individual institutions that might have very large mean residuals). No matter which approach is used, however, we are likely to find that college quality has a positive "effect" on achievement, even if there is in fact no effect. The reason for this is that error of measurement in the NMSQT causes us to underestimate the correlations of NMSQT with college quality and GRE scores, and thereby to "undercorrect" for initial NMSQT performance. Thus, even though we have statistically equated the student bodies entering each college in terms of their mean observed NMSQT scores, we have not equated them in terms of their mean true NMSQT scores: Since the adjusted "true" score is still positively correlated both with college quality and GRE scores, we should still expect to find a positive correlation between quality and GRE performance.

These artifacts can perhaps be better illustrated in terms of regression analysis. In simple linear regression, the slope of the regression line (regression coefficient) is a direct function of the correlation coefficient: $b_{yx} = r\frac{s_y}{s_x}$ (b and r, of course, will be identical if the two variances are equated). Thus, if _r_ is attenuated by error of measurement in x (NMSQT scores), then b_{yx} will also be attenuated. The net effect of error of measurement in our example, then, is to flatten the slope of the regression of GRE on NMSQT scores. Figure 3 shows the consequences of this phenomenon: If one flattens the



-------- Observed Regression (error of measurement in NMSQT)

________ True Regression (no error of measurement in NMSQT)

Figure 3. True and Observed Regression of GRE Scores on NMSQT Scores.

slope of the regression line (dotted line), he will tend to underestimate GRE for high values of NMSQT, and to overestimate GRE for low values of NMSQT. Therefore, if we attempt, statistically, to equate students entering different colleges by partialling out the effects of their initial ability (NMSQT scores) on GRE performance, the residuals for students with above-average NMSQT scores will be too large, and those for below-average students too small. Similarly, the mean residuals for students attending high-quality colleges will be spuriously large (because there are more high-ability than low-ability students at these institutions) and the mean residuals for students attending low-quality colleges will be spuriously small. The magnitude of these spuriously large residuals is a direct function of the amount of measurement error in the NMSQT.

The same problem occurs, of course, if we compute residual college "quality" scores: The residual environmental scores for students at high-quality colleges are spuriously high, and those for students at low-quality colleges spuriously low. In short, the net result of error of measurement in the NMSQT is that we undercorrect for initial differences in ability, thereby creating a spurious positive partial or part correlation between college quality and GRE performance.

A real danger in spurious "effects" of this sort is that they are likely to be believed by educators and policy-makers because they confirm existing theories about college impact. It is widely held, for example, that students get a "better education" in the "good" colleges. Indeed, this is one of the beliefs that attracts the brightest students to such colleges. Moreover, since most highly selective colleges manifest other traditional signs of prestige or quality (large libraries, distinguished faculty, competitive atmosphere, and so forth), the expectation that the student's intellectual development will prosper more in a high-quality college than in one of low quality is reinforced.

The same "believability" of results may be created by errors of measurement in many other types of input variables. Thus, we would expect students to become relatively more "liberal" if they attend colleges where the students are already highly liberal. We would expect a student with strong science interests to maintain these interests during college if he attend an institution where a high proportion of his fellow students also have strong science interests. The notion that students tend to change in the direction of their fellow students' dominant characteristics has been stated by Astin (1965a) and Astin and Panos (1969) as the theory of "progressive conformity" and by Feldman and Newcomb (1969) as the theory of "accentuation of initial differences." The point here is simply that, even if such theories are wrong, they will tend to be confirmed by virtue of error of measurement in the student input variables.

The fundamental importance of measurement error bias in studies of college impact indicates that some statistical proof of the bias be presented. For this purpose we can use the example of the study of the effects of college selectivity on GRE performance. The three basic variables can be designated with subscripts as follows:

0 = GRE score

1 = NMSQT score

2 = College Quality

Let us begin by assuming that (a) the correlations among these three variables are all nonzero and positive (which happens to be the case), but (b) there is no true effect of College Quality on GRE performance. In other words, let us assume that if it were possible to obtain error-free measurements of all three variables, the partial correlation between GRE and College Quality, holding constant the effects of NMSQT scores, would be zero. Using the familiar formula for a first-order partial correlation coefficient, the true partial correlation between quality and GRE performance would thus be:

$$(1)\ \text{(true scores)}\ r_{02\cdot 1} = \frac{r_{02} - r_{01}r_{12}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = 0.$$

Since the denominator of this formula is always nonzero and positive (except in the limiting case of a zero-order correlation of 1.0 between two of the variables), the numerator must be zero to satisfy the requirement that the partial coefficient be zero. Thus, our hypothetical situation of no true effect of college selectivity requires that:

(2) (true scores) $r_{02} - r_{01}r_{12} = 0.$

Our hypothesis about the biasing effects of measurement error is that the partial correlation based on the <u>observed</u> scores of these three variables will be <u>non</u>zero and positive:

(3) (observed scores) $r_{02} - r_{01}r_{12} > 0.$

Our proof of (3) will consist simply of showing what happens when error of measurement is introduced into (2). First we must be able to estimate the correlations among <u>true</u> scores as shown in (2). In order to do this we need to know what the <u>reliabilities</u> of the three variables (r_{00}, r_{11}, and r_{22}) are. For purposes of discussion we shall assume that all three reliabilities are imperfect but nonzero:

(4) $0. < r_{00} < 1.$

(5) $0. < r_{11} < 1.$

(6) $0. < r_{22} < 1.$

Since the true correlation between two variables is equal to the ratio between their observed correlation and the geometric mean of their reliabilities, formula (2) can now be expressed in terms of correlations among observed scores as follows:

(7) (observed scores) $$\frac{r_{02}}{\sqrt{r_{00}}\sqrt{r_{22}}} - \frac{r_{01}r_{12}}{\sqrt{r_{00}}\sqrt{r_{22}}\sqrt{r_{11}}} = 0.$$

or,

(8) $$\frac{1}{\sqrt{r_{00}}\sqrt{r_{22}}}\left(r_{02} - \frac{r_{01}r_{12}}{r_{11}}\right) = 0.$$

Note that, in order for the left side of the equation to equal zero, the two terms within the parenthesis must be an equality, i.e.:

$$(9)\quad r_{02} = \frac{r_{01}r_{12}}{r_{11}}$$

Now if we compute the partial correlation between GRE and College Quality using observed scores but make no correction for unreliability in our measures, only the right-side of (9) is affected, i.e., we omit the correction for unreliability and the denominator becomes 1.0. (Making no correction for unreliability is equivalent to assuming that there are no measurement errors and that the reliability of the variable is 1.0.) Since $r_{11} < 1.0$ from (5):

$$(10)\quad r_{01}r_{12} < \frac{r_{01}r_{12}}{r_{11}}$$

and

$$(11)\quad r_{02} > r_{01}r_{12} > 0.$$

Consequently, when no correction is made, the parenthetical term in (8) becomes nonzero and positive. Since the two reliabilities, r_{00} and r_{22}, are also nonzero and positive, the entire left side of equation (8) and, hence, the observed partial correlation between GRE and college selectivity becomes nonzero and positive. In short, failure to adjust for error of measurement in the NMSQT will lead to the conclusion that achievement is favorably affected by college quality <u>when there is no true effect</u>.

Several additional interesting conclusions can be deduced from formula (8):

1. The principal source of bias results from measurement error in the <u>partialled</u> variable (variable #1 in equation 8). Failure

to adjust for such bias can:

a. Create spurious "effects" when there are no true effects.

b. Exaggerate the magnitude of true effects.

c. Reverse the sign of the true effect.

2. The direction (sign) of the bias resulting from errors of measurement in the partialled variable depends on the signs of its correlations with the other variables (r_{01} and r_{12}).
3. Failure to adjust for measurement error in either the independent or dependent variable will only attenuate the observed partial correlation between the independent and dependent variable.

Several remedies can be employed to compensate for error of measurement in the input variables when multivariate techniques are used to evaluate college impact. A remedy for the two variable case (input and output) has been proposed by Tucker, Damerin, and Messick (1966). An appropriate generalization of their approach to the multivariate case would be to compute correlation matrices using the variance in the "true" (rather than observed) scores of all independent variables (student input as well as environmental). Since the true variance in a set of scores is simply the product of the observed variance and the reliability of the measure, it would be relatively simple to make such corrections if the reliabilities of all of the input and environmental variables were known. Estimates of reliability are usually available for psychometric devices, though not for the demographic and other types of questionnaire data that characterize so much of the research on college impact. Recognizing the need for this information, the research staff of ACE is currently engaged in an extensive empirical study to obtain estimates of error of measurement in questionnaire data (Boruch and Creager, 1970).

Environmental Measurement

In the early days of research on college impact, investigators were primarily concerned with the significance of "the college experience" in a generic sense. That is, they were little interested in differences among different types of colleges and, by implication, they dealt with how "going to college" compares with "not going to college." However, with the steadily growing proportions of young people who go on to some form of higher education and with many city and state systems moving toward open admissions, the question of the impact of college (versus no college) is becoming more an academic than a practical question. In short, with the greatly expanding higher educational opportunities and the extraordinary diversity among institutions, the question of the impact of college is increasingly coming to be one of the <u>comparative</u> impact of different types of college experiences.

Environmental measures in comparative studies of college impact function chiefly to provide a basis for interpreting any observed differential effects. The simplest form of environmental measurement is simply to compare the effects of one college with the effects of another. The environmental "measure" in such studies is merely a dichotomy: college A versus college B. While such studies may prove interesting to the persons immediately concerned with the institutions being compared, the crudeness of this environmental measurement greatly limits the generalizability of the findings beyond the two institutions. To illustrate this problem, let us assume that we are interested in comparing the differential impact of a given state university and a given private liberal arts college on the development of students. We find that the liberal arts college, relative to the state university, increases the student's aspirations to go on to graduate school after completing the baccalaureate degree. (For purposes of discussion, we shall assume that this is indeed a differential effect of the two colleges and not an artifact arising from our failure to control adequately differential inputs to the two institutions.) How can we explain this observed difference in the relative impact of the two institutions? Is it because their faculties differ in the encouragement they give to students to go on for advanced training, or is it because the stimulation provided by the student peer groups differs? Can the result be caused by more subtle institutional difference in living conditions, type of college town, or administrative practices? Perhaps the relative neglect of undergraduate instruction at the state university makes the student cynical about the importance of graduate education. Clearly, the observed differential effect of the two institutions is subject to a variety of interpretations.[11]

In this example, then, merely having available a large number of environmental measures on each of the two colleges does not resolve the interpretive dilemma, since there is no way to determine <u>which</u> of the various environmental attributes was responsible for the observed difference in institutional impact. Lacking any empirical way of choosing among the various measures, the investigator is forced to rely on a purely clinical or intuitive explanation of the observed effect.

The most obvious solution to this dilemma is to study simultaneously a much larger number of colleges with widely differing environmental characteristics. Ideally, the number would be large enough to permit reasonably reliable correlations using the institution as the unit of analysis. The relative contributions of the various environmental characteristics to the prediction of differential institutional impact would serve as an empirical basis for determining which environmental attributes are causally related to the student outputs under study.

Environmental measures are of two basic types: (1) The characteristics of the total institution (its size, selectivity, permissiveness, etc.) which can, in theory at least, affect all students at the institution and (2) special educational experiences _within_ the college (living in a particular dormitory, having a particular roommate, participating in an honors program, etc.) to which all students at a given institution are not exposed. This latter category comprises _within_-college environmental variables, whereas the former comprises _between_-college environmental variables.

Between-College Measures

Several instruments have been devised for measuring characteristics of college environments. In many ways, these instruments resemble personality inventories designed for assessing the traits of individuals; they include the _College Characteristics Index_ (CCI) (Pace and Stern, 1958; Stern, 1963); the _College and University Environmental Scales_ (CUES) (Pace, 1960, 1963); the _Environmental Assessment Technique_ (EAT) (Astin and Holland, 1961); and the _Inventory of College Activities_ (ICA) (Astin, 1968a). Reviews and discussions of these instruments have appeared elsewhere (Astin, 1968a; Menne, 1967). Each presents certain problems in inferring causation that merit some discussion.

These various instruments enbody three conceptually different approaches to the assessment of environmental characteristics: The "image" approach of the CCI and the CUES, the "student characteristics" approach of the EAT, and the "stimulus" approach of the ICA.

In the image approach, observers (usually students) are asked to report their impressions of what the college is like. The answers of all respondents at a particular institution are aggregated or averaged for each item, and the items are grouped into scales to form the environmental measures. Although the CCI and CUES use between 15 and 30 items per scale, data from the ICA (Astin, 1968a) indicate that highly reliable estimates of college "image" factors can be obtained with only two or three items. Apparently, when scale scores are averaged or aggregated across relatively large numbers of individuals, the advantages to be gained from basing each scale on a large number of items diminish.

The students characteristics approach is based essentially on an interpersonal theory of environmental influence. The objective is thus to assess the average or modal characteristics of the students at each institution. Although the EAT is based on only eight measures of the student body (size, ability, and six measures of personality), the possible number of relevant student characteristics is actually much larger.

The stimulus approach to measuring college environments was developed primarily because of certain interpretative difficulties connected with the image and student characteristic approaches. In the stimulus approach, the environment is seen as consisting of all of the stimuli that are capable of changing the student's sensory input. A "stimulus" is defined as "any behavior, event, or other observable characteristic of the institution capable of changing the student's sensory input, the existence or occurrance of which can be confirmed by independent observation" (Astin, 1968a, p. 5). This definition suggests that neither the college image nor the personal characteristics of the students satisfies the criterion of a potential stimulus. Thus, although the student's perception of his environment may influence his behavior toward his fellow students, his perception alone cannot function as a stimulus for others. Similarly, the student's intelligence, attitudes, values, and other personal characteristics do not constitute stimuli by this definition, even though such traits may be manifested in certain behaviors which can in turn can serve as stimuli for fellow students. The stimulus approach was thus developed in the belief that environmental measures based on such information would provide a better conceptual basis for interpreting causal relationships than either the image or student characteristics approaches.

A major difficulty presented by the image approach to measuring environmental characteristics is that the student's perception of his college can be influenced not only by what the college is really like but also by how it has influenced him. Thus, if a particular image factor is found to "affect" some student outcome, we can not be sure that we have adequately explained the observed effect, simply because the student's perceptions may have been influenced by the effect itself. This interpretative problem is well illustrated in a recent empirical study of differential college effects on the student's tendency to drop out of college (Astin and Panos, 1969). The environmental variable involved was Concern for the Individual Student, one of the eight "image" factors from the ICA (Astin, 1968a). It is defined by three image items: the percentage of students who rate the college environment as "warm" (scored positively), the percentage of students who agree with the statement that "most students are more like 'numbers in a book'" (scored negatively), and the percentage of students who agree with the statement, "I felt 'lost' when I first came to the campus" (scored negatively). This environmental measure correlated .38 with the mean residual (from input variables) percentage of students who stayed in college during the four years following matriculation. Although one might be tempted to interpret this finding in a direct causal sense (that is, it seems plausible that students are more likely to remain in college if their institution shows concern for them), it is also possible that the institution's score on this measure is the <u>result</u> rather than the cause of the student's persistence at an institution. That is, in responding to these college image questions, students who have already dropped out, or are about to, may be more inclined to report that their institution shows little concern for them than students who have made up their minds to stay in college. Of course, if the image factor has higher correlations with the outcome criterion than does any other environmental factor (which happens to be true in this particular case), then the investigator can more safely infer that the environmental measure is an antecedent rather than a consequence of the output variable under investigation.

A fourth category of between-institution environmental information -- one which has not as yet been used in developing an inventory to measure college characteristics -- is the structural and organizational characteristics of the institution. There are many such measures that might be used, both qualitative and quantitative. Among the more common qualitative measures are type of control, religious affiliation, type of curriculum, highest degree conferred, geographic region, sex, and race of the institution. Among the many possible quantitative measures are size of the institution, faculty-student ratio, tuition charges, endowment, operating budget, research funds, percentage of Ph.D.'s on the faculty, and library size. Such characteristics present certain interpretative difficulties because they are remote from the student and his development. From a practical standpoint, however, they are of particular importance, being more amenable to direct manipulation than are most of the measures that characterize the various environmental inventories. This fact suggests that we badly need to do research on the manner in which these structural administrative characteristics affect the college environment, and in turn, the development of the student. Creager and Sell (1969) have taken a step in this direction by developing a master institutional file, which contains measures of many structural and administrative attributes for all colleges in the population.

In one recent multi-institution longitudinal study, Astin and Panos (1969) compared image, personal characteristics, and stimulus measures in terms of their effectiveness in accounting for differential institutional impact on the undergraduate student's educational and career plans (Tables 32, 33, 55, and 56). In general, the stimulus measures accounted for larger proportions of the differential institutional effects than did either the personal characteristics or the image measures. Nevertheless, some outputs were most highly dependent on personal characteristics measures, and others appeared to be most highly dependent on college image measures.

This same study also compared the relative efficiency of continuous measures of environmental "traits" (ICA factors, for example) and discrete or "type" measures of college attributes (Protestant colleges, for example). Virtually all of the outputs were much more highly predictable from college trait measures, although, in a few cases, certain college type characteristics appeared to carry substantial weight. In particular, technological institutions and men's colleges appeared to have certain effects on student's career plans which could not be accounted for by the several trait measures used.

In short, the empirical evidence so far indicates that differential college impact, whatever the output being investigated, cannot be attributed to any single college environmental characteristic. Furthermore, while certain types of measures appear to be more effective in accounting for differential impact than others, no single class of measures derived so far seems to assess all the important institutional attributes. The lesson to be drawn from this is that investigators should use multiple measures of college characteristics rather than limiting their measures to a single class of institutional attributes. At the same time, further intensive research is needed to develop improved taxnomic procedures for measuring college environments.

<u>Within-College Measures</u>

Some educators object to the use of between-college measures, such as those discussed in the preceding section, on the grounds that a measure of the environment of the total institution may be a poor reflection of the environment actually encountered by individual students. Since there are unquestionably many distinct subenvironments within given institutions, especially within the complex universities, measures of the "total" institutional environment will confound these distinctions. One practical problem in using between-institution measures to describe subenvironments within institutions is to define the appropriate environmental subunits. This task may be relatively simple in many universities, where colleges or schools are well-defined, although the mere existence of such colleges or schools does not necessarily mean that they are functionally independent. In some universities, for example, the students attending the technical college have little or no contact with students or professors in any of the other colleges. On the other hand, in other universities, such students may live in dormitories and attend classes with students from a variety of other colleges. Nevertheless, if functionally independent subunits within institutions can be identified, there seems to be no reason why these units cannot be treated as separate "institutions" in the analysis of between-college environmental effects.

There are a great many within-college environmental experiences that cut across organizational subunits like colleges of schools. The methodological challenge to the researcher is to identify such experiences, devise an appropriate means for measuring them, and determine whether or not each of his subjects encountered these experiences during college. Below is just a partial list of the many types of within-college environmental experiences which can affect the student's development:

1. The characteristics of individual professors and individual courses.

2. His course of study.

3. The amount of time he spends at various activities (studying, outside reading, recreation, sleeping, etc.).

4. The type and amount of counseling or advisement he receives.

5. His participation in special educational programs (honors program, year abroad, Washington semester, undergraduate research participation, etc.).

6. His living arrangements (dormitory, fraternity house, commuting from home, private apartment).

7. Number and types of his roommates.[12]

8. His use of drugs (tranquilizers, barbiturates, hallucinogens, narcotics, etc.).

9. The type and amount of financial aid he receives.

10. The hours he works and the type of work he does.

11. His marital status and number of children.

12. The availability to him of a private automobile.

Information about most of these within-environment experiences can be obtained directly from the student by means of a follow-up questionnaire. However, self-reports present certain potential dangers, depending upon the nature of the experience. Those experiences that require relatively little interpretation, such as whether the student had a scholarship, seem to present few problems; however, reports of how many hours the student spent in outside reading, or ratings of his professors of roommates, may be systematically biased, thereby producing interpretive ambiguities similar to those described earlier for measures of the college "image." In theory, of course, no student's report of his own environmental experiences can be regarded as experimentally independent of data on his input and output characteristics. In practice, however, the bias resulting from this lack of independence is probably minimal, as long as the environmental experience being reported is relatively objective and not open to misinterpretation by the student.

A related problem concerns the "randomness" of the experience itself. The basic problem here is to determine the extent to which the student himself was directly responsible for his being exposed to the particular experience, or, in cases where exposure was determined by others, the extent to which their decision was based on a knowledge of the student's input characteristics. In some colleges, for example, the assignment of students to dormitories, classes, professors, and other experiences is virtually random or at least haphazard. In others, the student has almost complete control over where he lives, who his roommates are, and what courses he takes. The most difficult interpretative problems arise in the case of experiences over which the student has very direct control (drug use, for example).

This discussion of environmental measurement indicates that a stepwise analysis of college impact would successively control the effects of various independent variables on the output variable in the following logical sequence:

1. Main effects of student input variables.

2. Interaction effects among student input variables.

3. Main effects of <u>within</u>-college environmental variables.

4. Interaction effects among <u>within</u>-college variables.

5. Interaction effects between student input and <u>within</u>-college environmental variables.

6. Main effects of <u>between</u>-college environmental variables.

7. Interaction effects among <u>between</u>-college environmental variables.

8. Interaction effects between student input and <u>between</u>-college environmental variables.

9. Interaction effects between <u>within</u>-college and <u>between</u>-college environmental variables.

Of course, the higher-order interaction effects among student input, <u>within</u>-college and <u>between</u>-college variables can also be studied, although the possible number is so large that the investigator would ordinarily limit his research to those higher-order effects that test specific hypotheses.

Methods of Data Collection

One methodological question which has received very little attention is the technique used for collecting empirical data in studies of college effects. Unfortunately, logistical considerations often limit the techniques available to an investigator. One difficulty in multi-institution studies, for example, is that the conditions for collecting data may vary systematically from institution to institution, particularly if the tests or questionnaires are administered to groups of students. Important institutional biases may be introduced by variations in the instructions given, in the manner in which students' questions are answered, in the time alloted for completing the task, and in the physical surroundings where the task is carried out. The major problem with such biases is, of course, that they affect most of the students at a given institution and thereby introduce systematic error into the interinstitutional comparisons.

Serious systematic errors of this kind are probably more likely to occur with group follow-ups or post-testing than with initial group pretesting, since the general environmental conditions associated with freshman orientation and registration (assuming that this is the period during which the pretest is administered) are probably much the same at most institutions. Moreover, students are likely to be more cooperative about providing information and completing forms at this time than at any other. In addition, the persons who administer the forms and the other students who are completing them are likely to be anonymous individuals to the new students, wherever he enrolls. In the case of follow-up or post-testing, however, the stiuation is usually very different. Students may have to be assembled in some ad hoc fashion to complete the task, thus introducing biases with respect to the time and place of testing. Some students who are well along into their senior year may strongly resent being asked to spend time on such a task. If the testing is carried out in existing classes, important biases may be introduced either because the professor resents the intrusion on his class time or because of the classroom interaction that has already developed during the term.

The most serious problem with group follow-up testing at the institution is that it excludes dropouts and early graduates. As a consequence, generalizations concerning environmental influences are limited to what may be the least affected group of students. A more subtle problem is the effects of college on the student's tendency to drop out may be confounded with college "effects" on other outcomes. If dropout-proneness is correlated with changes in other student outcomes, then those institutions that encourage potential dropouts to stay in college will also appear to "affect" these other outcomes. One protection for the researcher here would be to see if these other "effects" hold up once he has controlled for the institution's dropout rate. In other words, he might use the institution's dropout rate as a kind of input or control variable.

Perhaps the most realistic alternative to group follow-up testing at the institution is the mailed questionnaire sent to the student's home. This technique permits the investigator to follow-up all or a random sample of all entering students, including those who have dropped out or transferred. In one sense, the self-administered questionnaire represents the most <u>un</u>standardized of all data collection techniques. The real methodological advantage here, of course, is that these extreme variations in the conditions of administration are <u>confounded</u> in the interinstitutional comparisons. The reduction in precision that results from this confounding is a small price to pay in order to eliminate the systematic biases that almost inevitably result from follow-ups carried out at the institution. It should be pointed out, however, that mailed questionnaires cannot be used for follow-up assessments that require proctoring (e.g., achievement testing).

The principal methodological limitation of the mailed questionnaire is that not all students will cooperate, and that those who do complete and return the questionnaire are likely to be a biased subsample of the total group to whom questionnaires are sent. Several solutions to this problem are possible. In the first place, it should be recognized that nonrespondent bias is likely to have a greater effect on marginal tabulations of data than on associational measures. In fact, some evidence (Astin, 1968a, Appendix B) suggests that interinstitutional intercorrelations of questionnaire items are virtually unaffected by nonrespondent bias. Whatever biases do exist, however, can be compensated for in certain ways. If pretest or input data are available on all subjects (as is usually the case), they can be utilized to develop compensatory weights to be applied to the respondents' data. The basic idea here is to give relatively more weight to those respondents who most closely resemble the nonrespondents in their input characteristics. The weights can be developed from actuarial tables (Astin and Panos, 1969) or from regression analyses (Astin, 1970). The major objective of the regression analysis is to produce a set of weights which, when applied to the input data of the respondents, yields marginal tabulations on all items which are identical to the original marginals based on all subjects (respondents and nonrespondents). While there is no guarantee that such differential weighting will compensate for all significant nonrespondent biases, it will unquestionably eliminate or at least reduce some of the bias.

A more familiar technique for dealing with nonrespondent bias is to conduct additional, more intensive, follow-ups of subsamples of nonrespondents by means of special delivery or registered mail, telegrams, telephone calls, or other methods. An examination of the data thus secured can provide estimates of the possible effects of nonrespondent bias in the larger sample. If the bias proves to be serious enough, the same intensive follow-up procedures should be applied to the total sample. This method can also be used to assess the impact of the differential weighting described above. If the differential weights tend to make the follow-up marginals more consistent with the data as revealed by the intensive follow-up, then it can be assumed that the weighting procedure is working. In fact, the weighted marginals may be even more valid than the data from an intensive follow-up in which responses are obtained from, say, 85 percent of the subsample. The investigator would have reason to believe that such was the case if the weighted and unweighted marginals tended to straddle the corresponding marginals from his intensive follow-up subsample. Whether the weighted marginals were plausible would, of course, have to be assessed in light of the likely (or even possible) change in the values of the marginals that would occur if a 100 percent response could be obtained.

There have been very few empirical studies of the effects of compensatory weights on data. As was mentioned earlier, such weighting is probably important, if not essential, in reporting the marginal tabulations of respondent data. It has been shown, for example, that those students who do respond to follow-ups, in comparison with those who do not, are brighter, achieve at a higher level, are more motivated, and have more highly educated parents (Astin and Panos, 1969; Astin, 1970). Consequently, unweighted follow-up marginals are almost sure to be biased with respect to any item having to do with either ability or SES. (These findings also suggest that investigators who intend to use mail follow-up techniques should routinely plan to collect student input information on ability, past achievement, and parental education.) Very little is known, however, about the effects of weighting on associational measures. Because of our current ignorance on this matter, it would seem that, at a minimum, investigators should consider repeating some of their causal analyses using both weighted as well as unweighted follow-up data.

Although the systematic errors that result from group administration of instruments may seriously bias the analyses of comparative institutional impact, the investigator has at his disposal several techniques for detecting such biases. Obviously, he must make sure that the procedure used by the institution is thoroughly documented. When he finds that certain individual institutions show especially pronounced "effects" that cannot be accounted for by measurable environmental attributes, he should suspect that biases are operating. (To explore this question, he would need a large number of institutions and a comprehensive set of environmental measures.) The investigator will also have reason to suppose that systematic biases are present if the institution in question shows peculiarly large "effects" on highly subjective or judgmental outcomes. Presumably, the student's report of relatively factual outcomes, such as his final field of study of his marital status, is not as likely to be affected by situational factors as is his report of, say, his personal values or attitudes. In particular, the students' ratings or subjective judgments of his college are probably most sensitive to variations in the institutional circumstances under which these judgments are obtained. Routine inclusion of a few such items in every group-administered questionnaire may serve as a good check on the possibility of systematic biases.

Resolving Some Inferential Dilemmas

No matter how elegant his design, the investigator can never be absolutely sure that he has isolated "true" college effects. But there are certain situations in which he can have a good deal of confidence that his data are indeed revealing environmental influences.

One situation which justifies a high degree of confidence is that in which _the environmental variable is uncorrelated with the input variables_. For example, one college environmental characteristic that has no relationship, or only a very low one, with most student input variables is institutional _size_ (Astin, 1965b). In other words, students who go to large institutions differ very little from those who go to small ones. Consequently, the observed effect of size on some student output measure is almost certainly not just an artifact of the researcher's failure to control input differences.

A related situation arises _when the correlation between environment and output is substantially higher than the correlation between environment and input_. Ideally, one would like to see the correlation between environment and output increase as differential inputs are controlled. (The more usual situation, of course, is that the environment-output relationship shrinks consistently as input variables are successively controlled.) Thus, if the environment-output correlation increases, or at least holds its own, as input variables are controlled, the investigator can be reasonably confident that his observed environmental effect is a true one. However, if the correlation between environment and output diminishes consistently as input variables are controlled until only a small relationship remains, there is a very real possibility that if one or two other input variables had been included in the analyses, the relationship would have disappeared altogether.

Perhaps the strongest evidence for true causation exists when the director (sign) of a particular environmental effect is the _opposite_ of the zero-order correlation between that environmental variable and the output measure. Although rare, this reversal in sign has been observed in at least one study (Astin and Panos, 1969). In this study, Cohesiveness, an ICA environmental measure reflecting primarily the proportion of students who report having many close friends among their fellow students, was shown to have a positive effect on the student's chances of staying in college. The zero-order correlation between Cohesiveness and the percentage of students remaining in college for four years, however, was _negative_ (r = -.13). When differential student input variables were controlled, this partial correlation reversed sign (to +.25). The explanation of this apparent paradox is that students who go to highly cohesive institutions are, on the average, more dropout-prone than are the students who go to the less cohesive institutions. Consequently, the positive relationship between the dropout-proneness of entering freshman classes and the Cohesiveness of institutional environments masks the negative _effect_ of Cohesiveness on the individual student's chances of dropping out.

Research data that reveal significant _interaction_ effects represent another situation where causal inferences can be made with more than usual confidence. Consider, for example, a situation where a measure of, say, the peer environment is found to have a significant main effect on some outcome. If this apparent effect is indeed a true one rather than an artifact, then we would expect to find that the effect is stronger among resident students -- who would have more contact with their fellow students and thus would be more affected by their characteristics -- than among commuters. Therefore, if it can be shown that there is a significant interaction effect involving resident-commuter status and the particular peer environmental measure in question, then the conclusion that the environmental attribute is causally related to the outcome is strenghtened. By the same reasoning, we should not expect to find such interactions with measures of, say, the classroom environment, since both residents and commuters presumably have equal exposure to such environmental factors by virtue of attending classes.

Similar checks on the validity of causal inferences can be made by examining many other types of interaction effects. For example, extroverts or gregarious students are presumably more susceptible to the effects of peer factors than are introverted or shy students. To take another example, the magnitude of a particular effect should increase the longer the student is at the college. The point is simply that, for many of the apparent environmental effects that may be observed in longitudinal studies, it is possible to hypothesize the existence of certain interaction effects which, if subsequently confirmed by additional analyses, would lend support to the assumption that the relationship is indeed a causal one.

Summary

The purpose of this paper has been to review some of the major methodological problems in the design of studies of college impact.

To facilitate the discussion of design problems, one may view the question of college impact in terms of three components: student outputs, student inputs, and environmental characteristics. Any problem in the design of college impact studies can be seen in terms of the relationships between these <u>three</u> components.

The major goal of college impact studies is to minimize three kinds of inferential error: type I and type II errors (the traditional inferential errors of experimental design), and type III errors, which are defined as inferential statements which simultaneously involve both type I and type II errors. Type III errors are possible in college impact research primarily because of the highly nonrandom distribution of students among institutions.

Much of the previous research on college impact has resulted in ambiguous findings primarily because at least one of the three informational components was missing. The single-institution study, through input and output information, indicates how the student changes during college, but it provides no information bearing directly on environmental impact. The multi-institution cross-sectional study provides information on the relationship between environments and outputs, but it is highly susceptible to type I and type III errors unless student input data are also collected.

The most definitive information about college impact is obtained from multi-institution longitudinal studies in which data on student inputs, student outputs, and environmental characteristics are obtained.

Such data can be analyzed by a variety of "quasi-experimental" designs (path analysis, for example), although step-wise linear multiple regression analysis is perhaps the most flexible and versatile method, particularly if the regression is carried out in separate "stages" dictated by the logic of the college impact process.

The variance in student output in the multi-institution studies can be assigned to four sources: error, variance uniquely attributable to input variables, variance uniquely attributable to environmental variables, and confounded variance. Variance uniquely attributable to input variables can be defined as the squared multiple part correlation between the output measure and the residual input measure. Variance uniquely attributable to environmental sources can be defined as the squared multiple part correlation between the output measure and the residual environmental measures. Confounded variance is defined as the remainder of the total predictable output variance (that is, the final $\underline{R}^2$ minus the two squared part correlations). The two part correlations can be used as "lower-bounds" estimates of the total output variance attributable to a particular source, whereas the part correlation plus the confounded variance can be used as an "upper-bounds" estimate. Additional part correlations and confounded variance estimates can be obtained for interaction effects, if desired.

One of the most serious sources of potential bias in college effects studies, regardless of the method of analysis used, are errors of measurement in the input variables. Unless corrections are made for such errors, the investigator runs the risk of finding spurious college "effects." Such spurious effects tend to be highly believable, in that they ordinarily support the most plausible theory of how students are affected by their colleges. Since appropriate adjustments for such measurement error require a knowledge of the reliability of each input variable, researchers engaged in studies of college impact should consider routinely collecting such reliability information on each instrument that they use.

The generalizability and usefulness of information about college impact depends very highly on the number and kinds of the environmental measurements used. The principal function of environmental measurement in research on college impact is to provide an interpretive frame of reference for any significant effects that might be observed. Although the most popular approaches to environmental measurement have been based on student perceptions of the environment, such measures present interpretive difficulties in view of the possibility that the effect itself may have influenced the student's perception of his institution. One possible solution to this problem is to develop environmental measures based on directly observable events rather than on perceptions.

In spite of the many methodological and logical problems inherent in research on college impact, several checks and precautions are available to the investigator that will reduce his chances of commiting inferential errors.

Footnotes

[1]This study was supported by grants from the National Institute of Mental Health and the National Science Foundation, and by general funds from the American Council on Education. The author is indebted to Alan E. Bayer, Robert F. Boruch, and David E. Drew for their criticisms of an earlier draft of this paper, and especially to John A. Creager for his suggestions concerning the section on measurement error. Portions of this paper were presented at the Annual Meeting of the American Educational Research Association, Minneapolis, 1970.

[2]This model, which was originally presented in Astin (1965a), has been adapted for a program of research in higher education (Astin and Panos, 1966) and for a more general model of educational evaluation (Astin and Panos, 1970).

[3]In statistical terms, "opposite" means that there has been an error of direction or sign in rejecting a two-tailed null hypothesis. A similar idea has been proposed earlier by Kaiser (1960).

[4]See Astin, Panos, and Creager (1966); Panos, Astin and Creager (1967); Creager, Astin, Boruch, and Bayer (1968); and Creager, Astin, Bayer, Boruch, and Drew (1969).

[5]Part correlations, which are not described in most textbooks on statistics, may be unfamiliar to some readers. Simple zero-order correlations involve two unadjusted variables. Partial correlations involve two residual variables (the residuals having been calculated from a third variable or set of variables). Part correlations involve one unadjusted variable and one residual variable.

[6]Actuarial analysis also bears certain similarities to analysis of variance. A major difference, of course, is that the cells (treatments) are formed

in actuarial analysis on the basis of a knowledge of the dependent (output) variable, rather than a priori.

[7]The first method, where student inputs are controlled using the student as the unit of analysis, would not entirely partial out environmental effects under these circumstances, since the within-college regressions would not fall along the common or pooled regression line. Thus, the residuals in colleges with initially high mean input scores would tend to be positive, whereas the residuals in colleges with relatively low mean input scores would tend to be negative.

[8]A suppressor variable is one whose addition to a set of independent variables increases the beta weight associated with one or more of the variables in the set.

[9]It should be noted, however, that even these part correlations are not independent (Creager, 1970). For a fuller treatment of the problem of collinearity, see Creager and Boruch (1970).

[10]There may, of course, be interactions among input variables or among environmental variables; these types of interaction effects can be dealt with by rescoring the variables involved.

[11]These interpretive ambiguities inherent in "comparative" studies of this type have been discussed at length by Cronbach (1963).

[12]An interesting approach to defining any individual student's peer environment has been described by Rossi (1966) and employed by Wallace (1963). Briefly, this technique develops environmental measures from the aggregated characteristics only of those fellow students who are close friends or associates of the student in question.

References

Astin, A. W. "Productivity" of Undergraduate Institutions. Science, 1962, 136, 129-135.

__________. Differential College Effects on the Motivation of Talented Students To Obtain the Ph.D. Degree. Journal of Educational Psychology, 1963(a), 54, 63-71.

__________. Undergraduate Institutions and the Production of Scientists. Science, 1963(b), 141, 334-338.

__________. Effects of Different College Environments on the Vocational Choices of High Aptitude Students. Journal of Counseling Psychology, 1965(a), 12, 28-34.

__________. Who Goes Where To College? Chicago, Illinois: Science Research Associates, 1965(b).

__________. The College Environment. Washington, D. C.: American Council on Education, 1968(a).

__________. Undergraduate Achievement and Institutional "Excellence." Science, 1968(b), 161, 661-668.

__________. Manual for the ICA. In press, 1970.

__________., and Holland, J. L. The Environmental Assessment Technique: A Way to Measure College Environments. Journal of Educational Psychology, 1961, 52, 308-316.

__________., and Panos, R. J. A National Research Data Bank for Higher Education. Educational Record, 1966, 47, 5-17.

__________., and __________. The Educational and Vocational Development of College Students. Washington, D. C.: American Council on Education, 1969.

__________., and ____________. Evaluation of Educational Programs. Chapter 19 in forthcoming revision of Educational Measurement (R. L. Thorndike, Ed.), Washington, D.C.: American Council on Education, 1970.

Bayer, A. E., and Astin, A. W. Violence and Disruption on the U. S. Campus, 1968-1969. Educational Record, 1969, 50, 337-350.

Boruch, R. F., and Creager, J. A. Investigation of Measurement Error in Educational Survey Research, 1970. (Proposal submitted to the Office of Education, January 16, 1970.)

Bottenberg, R. A., and Ward, J. H., Jr. Applied Multiple Linear Regression. Technical Documentary Report PRL-TDR-63-6. Lackland Air Force Base, Texas: 6570th Personnel Research Laboratory, Aerospace Medical Division, March 1963.

Campbell, D. T., and Stanley, J. D. Experimental and Quasi-experimental Designs for Research on Teaching. Handbook for Research on Teaching. (Edited by N. L. Gage.) Chicago: Rand McNally, 1963.

Creager, J. A. On Methods for Analysis and Interpretation of Input and Treatment Effects on Education Outcomes. American Council on Education. (Mimeo.)

__________., Astin, A. W., Bayer, A. E., and Boruch, R. F. National Norms for Entering College Freshmen - Fall 1968. ACE Research Report, 1968, 3, No. 1.

__________., __________., __________., __________., and Drew, D. E. National Norms for Entering College Freshmen - Fall 1969. ACE Research Report, 1969, 4, No. 7.

__________., and Boruch, R. F. Orthogonal Analysis of Linear Composite Variance. Proceedings, 77th Annual Convention, American Psychological Association, 1970, 113-114.

__________., and Sell, C. L. The Institutional Domain of Higher Education: A Characteristics File. _ACE Research Report_, 1968, 4, No. 6.

Cronbach, L. J. Course Improvement Through Evaluation. _Teachers College Record_, 1963, 64, 672-683.

Darlington, R. B. Multiple Regression in Psychological Research and Practice. _Psychological Bulletin_, 1968, 69, 161-182.

Duncan, O. D. Path Analysis: Sociological Examples. _The American Journal of Sociology_, 1966, 72, No. 1, 1-16.

Feldman, K. A., and Newcomb, T. M. _The Impact of College on Students_. San Francisco: Jossey-Bass, Inc., 1969.

Holland, J. L. A Personality Inventory Employing Occupational Titles. _Journal of Applied Psychology_, 1957, 42, 336-342.

__________. Determinants of College Choice. _College and University_, 1959, 35, 11-28.

Kaiser, H. F. Directional Statistical Hypotheses. _Psychological Review_, 1960, 67, 160-167.

Knapp, R. H., and Goodrich, H. B. _Origins of American Scientists: A Study Made Under the Direction of a Committee of the Faculty of Wesleyan University_. Chicago: University of Chicago Press, 1952.

__________., and Greenbaum, J. H. _The Younger American Scholar: His Collegiate Origins_. Chicago: University of Chicago Press; and Middletown, Conn.: Wesleyan University Press, 1953.

Menne, J. W. Techniques for Evaluating the College Environment. _Journal of Educational Measurement_, 1967, 4, 219-225.

Pace, C. R. Five College Environments. _College Board Review_, 1960, 41, 24-28.

__________. _College and University Environmental Scales_. Princeton, N. J.: Educational Testing Service, 1963.

Panos, R. J., Astin, A. W., and Creager, J. A. National Norms for Entering College Freshmen - Fall 1967. ACE Research Report, 1967, 2, No. 7.

Richards, J. M., Jr. A Simple Analytic Model for College Effects. School Review, Winter 1966, 74, 380-392.

Rossi, P. H. Research Strategies in Measuring Peer Group Influence. In T. M. Newcomb and E. K. Wilson (Eds.), College Peer Groups: Problems and Prospects for Research. Chicago: Aldine, 1966.

Stern, G. G. Scoring Instructions and College Norms for the Activities Index and the College Characteristics Index. Syracuse, N. Y.: Psychological Research Center, 1963.

Tucker, L. R., Damarin, F., and Messick, S. A Base-free Measure of Change. Psychometrika, 31, 1966, 457-473.

Wallace, W. L. Peer Groups and Student Achievement: The College Campus and Its Students. Report No. 91. Chicago: National Opinion Research Center, University of Chicago, 1963.

Werts, C. E. Path Analysis: Testimonial of a Proselyte. The American Journal of Sociology, 1968, 73, No. 4, 509-512.

__________., and Watley, D. J. A Student's Dilemma: Big Fish--Little Pond or Little Fish--Big Pond. Journal of Counseling Psychology, 1969, 16, 14-19.