APPENDIX:

Methodological and Analytical Issues in Assessing the Influence of College

This appendix is designed to provide an introduction to some of the more important methodological and analytical issues in estimating the influence of college. It covers the following basic topics: the problem of estimating unique or net effects, attributing student change to college attendance, attributing student change to the type of college attended or to different experiences within the same institution, analytical approaches useful in assessing the impact of college (that is, residual scores and partial correlation), different multiple correlation and regression procedures (including commonality analysis, hierarchical and stepwise analysis, and causal modeling), issues in the appropriate unit of analysis, the use of change scores, and conditional versus general effects.

The Problem of Estimating Unique or Net Effects

One of the most basic and persistent questions in educational research has been how one determines the extent to which student change or development can be attributed to the educational experience itself and not to other factors or competing influences. We will refer to this as the unique or net effects question, although more recently it has been referred to by other names, such as the value-added question (Astin, 1985a, 1985b; Ewell, 1985a, 1985b; Pascarella, 1986b). At the very heart of this question is an attempt to separate that part of student change that is caused by the particular educational experience under investigation from the part that is due to other influences, such as student background abilities or normal maturation over time. It has been axiomatic in the educational research community that the most valid approach for estimating the causal link between two variables and thus the net effect of one on the other is through the random assignment of individual subjects (or an alternatively appropriate unit of analysis) to experimental and control groups (for example, Astin, 1970c, 1970d; Borg & Gall, 1983; Campbell, 1967; Campbell & Stanley, 1963; Cochran & Cox, 1957; Cook & Campbell, 1979; Cronbach, 1982; Kerlinger & Pedhazur, 1973; Linn, 1986; Linn & Werts, 1977; Rubin, 1974; Stanley, 1967).

Unfortunately, the necessary conditions for a true or randomized experiment are extremely difficult to obtain in actual field settings where self-selection rather than random assignment is the rule. This is no less a problem in research on the influence of college than it is in educational research generally. Indeed, a basic and perhaps *the* basic problem in assessing the unique influence of college on students is the issue of student self-selection or recruitment. An overwhelming body of evidence suggests that characteristics are not randomly distributed to college and noncollege groups (for example, Anderson, Bowman, & Tinto, 1972; Astin, 1975c; Bachman & O'Malley, 1986; Blau & Duncan, 1967; Bowles & Gintis, 1976; Braddock, 1980; Brim, Glass, Neulinger, & Firestone, 1969; Peng, Bailey, & Eckland, 1977; Sewell, 1971; Sewell & Shah, 1967, 1968b; Taubman & Wales, 1973; Thomas, Alexander, & Eckland, 1979; Trent & Medsker, 1968; Wolfle, 1985b), to different types of colleges (for example, Alexander & Eckland, 1977; Astin, 1961, 1962, 1965c, 1982; Feldman & Newcomb, 1969; Hearn, 1984; Karabel & Astin, 1975; Medsker & Trent, 1972; Bohrnstedt, 1967; Burns, 1974; Smart & Pascarella, 1986b), or to different academic and social experiences within the same institution (for example, Astin, 1971b, 1977a, 1984; Bohrnstedt, 1967; Burns, 1974; Pascarella, 1980, 1985b; Pascarella, Smart, Ethington, & Nettles, 1987). This means that in estimating the influence of college on student development, one is confronted with the particularly knotty problem of separating the influence due to collegiate experiences alone from that due to the particular characteristics of the individuals participating in those experiences.

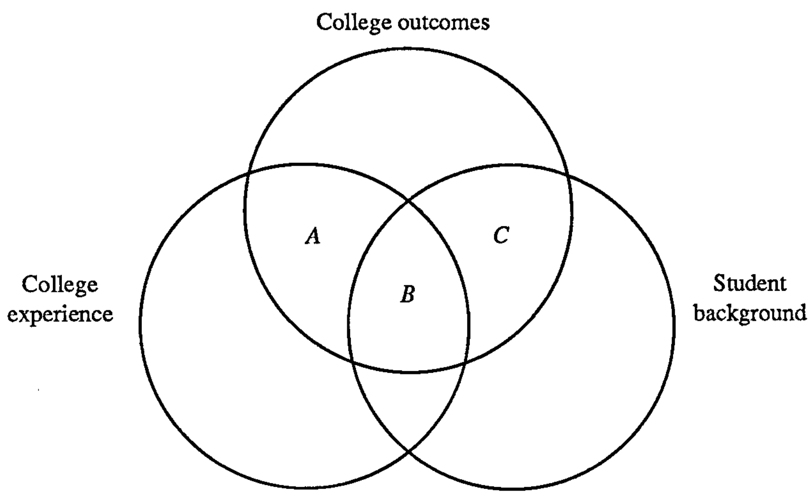
It is typically the case that student background characteristics (academic aptitude, prior achievement, family socioeconomic status, aspirations, personality orientations, and the like) are not merely the best predictors of many of the outcomes associated with college; they are also a major determinant of whether or not one attends college and, if so, the type of college attended and the extent and quality of involvement in different academic and social experiences during college. Because individual student background characteristics influence both categories of variables (the collegiate experience and outcome measures), they satisfy the requirements for the classical definition of a confounding variable—that is, a variable so associated with both the independent (collegiate experience) and dependent (outcome) variables that it may be confused with the effect of the independent variable (for example, Cohen & Cohen, 1975; Cronbach, 1982; Kerlinger, 1979; Kerlinger & Pedhazur, 1973; Pedhazur, 1982). In short, student background characteristics may masquerade as college effects. Another way of saying this is that the associations between college experience and outcomes may be spurious or noncausal because both are dependent upon mutually antecedent causes such as individual student background characteristics (Linn & Werts, 1969).

This problem is demonstrated graphically in Figure A.1. As Figure A.1 shows, a substantial part of the relationship between college experience and college outcomes (that is, the area defined by $A + B$) is confounded by the joint relationship between student background and both college experience and college outcomes (area B). Determining or estimating the unique or net influence of college experience on college outcomes (that is, $[A + B] - B$) is a nontrivial analytical problem, and we now turn to some ways in which the issue has been addressed.

Attributing Student Change to College Attendance

One fairly common methodological approach in estimating the influence of college attendance (versus nonattendance) on students has been to treat those attending or graduating from college as the experimental or "treated" group while designating those whose formal education stops with secondary school as the natural control group (for example, Colby, Kohlberg, Gibbs, & Lieberman, 1983; Hyman & Wright, 1979; Hyman, Wright, & Reed, 1975; Plant, 1962; Rest & Thoma, 1985; Telford & Plant, 1963; Trent & Medsker, 1968; Wolfle & Robertshaw, 1982). The problem, of course, is that the control group (those not attending college) may be so fundamentally different in important individual characteristics from the treated group (those exposed to college) that simple comparison of outcome differences is often meaningless. Thus, for example, if given a mea-

Figure A.1. Schematic View of Variable Relationships in Estimating College Effects.



sure of vocabulary or verbal learning, college graduates would likely score higher than high school graduates. A substantial part of this difference, however, would probably be due to the fact that when they graduated from high school both groups also differed in the same direction in verbal ability. Indeed, to some extent it is higher levels of verbal ability that led the college group to be admitted to college in the first place.

One way of dealing with this problem is to attempt to match or equate college and noncollege groups on important background characteristics, that is, to select for the control group those high school–only subjects who are similar in verbal ability to their counterparts who attended college (King, Kitchener, & Wood, 1985; Kitchener & King, 1981; Pascarella, 1989). Though this approach is workable when subjects are being matched on a small number of variables (for example, one to three), it can quickly become impractical if one attempts to match subjects on an extensive number of variables. Another, more common approach has been to use analytical procedures such as partial correlations, regression analysis, or analysis of covariance to "statistically equate" the comparison groups (for example, Robertshaw & Wolfle, 1982; Wolfle, 1983). What these approaches typically accomplish is to statistically remove (or "partial out" or "covary out") the confounded part of the relationship between the independent and dependent variables. The unconfounded "part" of the correlation that remains is often interpreted as the unique or net association between the independent and dependent variables (Cohen & Cohen, 1975).

Such statistical procedures are discussed and expanded upon later in this appendix. It is important to point out at this juncture, however, that these statistical procedures are not a totally adequate substitute for random assignment of subjects to treatments (Elashoff, 1969; Linn & Slinde, 1977; Lord, 1967, 1969). Indeed, as suggested by Lord (1967, p. 305), "There simply is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled preexisting differences among groups." Moreover, statistical control (like control by matching) is further limited to those variables for which data are available. Other important but unmeasured variables may remain uncontrolled and therefore lead to problems in interpretation and attribution. Despite these significant limitations, the vast preponderance of research on the influence of college on students has had to rely on less than ideal statistical controls in estimating the net effects of college experiences (Astin, 1970a, 1970b; Feldman, 1969, 1971b; Pascarella, 1982; Wolfle, 1985a).

Even if one is able to match groups or statistically equate them on important background characteristics, the college versus noncollege comparison may still be confounded by differences in motivational factors that lead some individuals to invest in college while others go from high school directly to the labor force. This is a subtle and therefore pernicious threat to the validity of the findings that has been referred to as the "interaction of selection and change" or "the interaction of selection and the dependent variable" (Campbell & Stanley, 1963). What it suggests is that the complex set of motivations underlying a person's choice of a particular experience (such as attending college) may itself be an important determinant of many of the outcomes typically associated with the experience.

For example, consider two high school graduates of equal ability and family socioeconomic position, with one choosing college and the other not. The college graduate may eventually have a higher level of income not just because a college degree provides entry into higher-salaried professions but also because of the higher level of ambition, drive, or need to achieve that led him or her to choose to attend college in the first place. Underlying ambitions, drives, and achievement needs may be particularly difficult to assess; and if they cannot be assessed, then confounding effects cannot be statistically removed. Often this leads one to attribute to the collegiate experience an effect that is more appropriately attributed to the individual characteristics of those who attend college.

With few exceptions (for example, Stanley, 1967), there may be no naturally occurring and easily accessible control group of noncollege individuals. This has apparently been recognized by researchers who have attempted to assess the effects of college on students without employing control groups of individuals not exposed to college. One typical design is the simple pretest-posttest longitudinal design in which the same panel or sample of students is followed over a specified period of time, such as freshman to senior year, and measured on the same instrument. The students are essentially their own control group, and the difference between mean freshman and senior scores on some measure of interest (moral reasoning, reflective judgment, or critical thinking, for instance) is used as an estimate of the effect of college. Unfortunately, even if one assumes the reasonable reliability of change scores (an unsafe assumption; see Cronbach & Furby, 1970), such mean changes may reflect not only the influence of college but also the effects of confounding noncollege influences such as external events (history), the practice effect (taking the same instrument or test twice), and possibly even regression (to-the-mean) artifacts if the group is extremely low or high on the first testing. In many instances, however, the most troublesome confounding variable associated with simple longitudinal panel designs having no control group is that of age or maturation. From the perspective of many developmental theorists, a considerable portion of the maturing that occurs coincidentally with college attendance may be explainable by the fact that students grow older as they attend college (for example, Erikson, 1968; Kohlberg, 1969; Loevinger, 1976; Perry, 1970). This, of course, implies that average freshman-to-senior changes may well overestimate the net influence of college attendance (McMillan, 1986).

Another problem with this design is mortality or attrition in student responses from freshman to senior year. Students willing to be in a study or to respond to a questionnaire instrument as freshmen may not always be as willing to do so as seniors. Since it is necessary to have pre- and post-measures for each student, those for whom freshman-to-senior comparisons are possible may not be representative of the institutional population to whom the results are to be inferred. This, of course, may speak more to the generalizability of the findings (external validity) than to the extent to which the differences can be attributed to the college experience and not alternative or competing hypotheses (internal validity).

In an effort to gain some control over the confounding effects of student maturation, a number of cross-sectional (cohort) designs have been suggested as an alternative to the longitudinal pretest-posttest panel design without control group. Consider the cross-sectional design where freshmen are compared with seniors on a measure of critical thinking. The freshmen, who have not been exposed to college, act as a control group for the seniors, who have theoretically benefited from four years of exposure to it. (To better reflect the entire college experience, the measure of critical thinking might be given to freshmen upon enrollment in college and seniors in the final semester or quarter of their senior year.)

Since this design is comparing different cohorts on the same measure, the potential confounding influences of regression artifacts or of being assessed twice on the same measure are largely controlled. There is still the problem of age differences between freshmen and seniors. In a cross-sectional design, however, it is at least possible to adjust statistically for the effects of age. Other weaknesses in this design remain, of course. First, because of attrition, seniors may represent a more selective population in terms of academic ability and related traits than freshmen, although this may vary with the patterns and determinants of attrition at different colleges. Second, there is the possibility that differential recruitment or admission criteria may be used for the seniors versus the current freshmen. (For example, if the college used a more stringent set of standards for admitting the senior cohort than it did in admitting the current freshmen, the former might represent a more academically select and motivated group than the latter.) Thus, to increase the validity of the findings yielded by such cross-sectional designs, one may need to adjust statistically for aptitude or prior achievement as well as for age (Mentkowski & Strait, 1983; Whitla, 1978).

Another cross-sectional design that attempts to disaggregate the effects of college from the effects of maturation without having to depend on statistical adjustments for age is suggested by Goulet (1975) and exemplified in the work of Strange (1978). This design requires that nonoverlapping cohorts of different ages be assessed on the dependent measure of interest at different points in their college career. For example, traditional-age freshmen (age eighteen), nontraditional-age freshmen (for example, age twenty-two), traditional-age seniors (age twenty-two), and nontraditional-age seniors (for example, age twenty-six) might all be administered the measure of critical thinking. The effects of maturation independent of college could be obtained by comparing average scores of samples of different ages but with the same amount of exposure to college (for example, freshmen age eighteen with freshmen age twenty-two and seniors age twenty-two with seniors age twenty-six). Conversely, the influence of college independent of age could be estimated by comparing average scores of samples of the same age but with different levels of exposure to college (for example, freshmen age twenty-two and seniors age twenty-two). While this design is quite powerful in detecting and holding constant the effects of age, it does not guarantee that the estimated differences associated with differential exposure to college can, in fact, be attributed to the net influence of college. Other factors, such as differential selectivity or recruitment effects between freshmen and senior samples, may still introduce uncontrolled bias in the results.

Attributing Student Change to Type of College Attended or Different Experiences Within the Same Institution

The design and analysis issues that confront researchers in their attempt to separate the influence of type of college attended or specific collegiate experiences from the confounding effects of individual subject differences are essentially the same as those in attempts to separate the effects of college attendance (versus nonattendance) from individual differences. Specifically, different types of colleges tend to recruit and enroll different kinds of students. Consequently, in comparing student outcomes associated with attendance at different kinds of institutions, it is difficult to separate recruitment effects from socialization effects; the former are the result of differential selection, while the latter refer to the actual effects of attending different colleges (Chickering, 1971b; Hauser, 1971; Kamens, 1971; Withey, 1971). Similarly, because different types of students tend to have different levels of involvement in college (for example, Astin, 1984; Pace, 1984) and to become involved in different types of academic and social experiences during college (for example, Apostal, 1970; Astin, 1977a; Astin & Panos, 1969), separating the influence of a specific set of college experiences from the influence of individual student differences is a nontrivial matter. Thus, for example, obtaining a valid estimate of the net influence of academic major on formal (Piagetian) reasoning is extremely difficult if one ignores the likelihood that students choosing different academic majors may differ substantially in level of formal reasoning to begin with.

Analytical Approaches to Assessing the Influence of College

A number of different analytical models have been used in estimating the influence of college on student development. While these methods differ somewhat in emphasis, they have in common a concern with separating the influence of college or different collegiate experiences (socialization) from the effects of individual student differences and self-selection (recruitment). We will discuss each of these approaches separately. Since our major purpose is didactic, we will emphasize intuitive and visual or geometric explanations rather than mathematical ones. We will also illustrate our discussion with an example based on plausible, though fictitious, data. We hope that this approach will aid the general reader of studies on the influence of college in better understanding and evaluating the meaning of results from different analyses.

Example Data

For our illustrative example, let's assume that we are interested in studying the influence of student-faculty informal contact on the development of intellectual orientation during the freshman year. To do so, we plan and conduct a longitudinal panel investigation in which we collect data at three points from the same sample of 1,000 students attending ten institutions that are comparable in student selectivity but differ substantially in enrollment size. At the beginning of the freshman year or prior to enrollment, all members of the sample are tested on a measure of intellectual orientation (IO_1). During January and February of the freshman year, a questionnaire is sent to members of the sample to obtain a measure of the frequency of their informal, nonclassroom contact with faculty to discuss intellectual or course-related matters (*IWF*). Finally, in May of the freshman year, we again test members of the sample on the measure of intellectual orientation (IO_2). Since we are also interested in the effects of institutional size or enrollment (*SIZE*), we collect data on this from institutional records and assign the appropriate institutional value to each student's data file.

Thus, our final set of data would consist of four measures for each student:

1. Precollege intellectual orientation (IO_1)
2. The size of the institution attended (*SIZE*)
3. Frequency of informal contact with faculty (*IWF*)
4. End-of-freshman-year intellectual orientation (IO_2)

Clearly, if this were an actual study, we might also be interested in other student preenrollment traits or other measures of institutional characteristics and freshman-year experiences. For didactic simplicity, however, we will limit our example data to the four variables noted.

The correlations among the four variables in our fictitious study are shown in Table A.1. As can be seen, the size of the institution attended, informal contact with faculty, and postfreshman intellectual orientation are

Table A.1. Matrix of Correlations.

Variable	*M*	*SD*	*SIZE*	*IWF*	IO_2
1. Precollege intellectual orientation (IO_1)	40.50	6.85	−.08	.13	.60
2. Institutional size (*SIZE*)	9,500	2.45		−.24	−.15
3. Informal interaction with faculty (*IWF*)	6.83	2.42			.33
4. Postfreshman intellectual orientation (IO_2)	51.53	7.16			

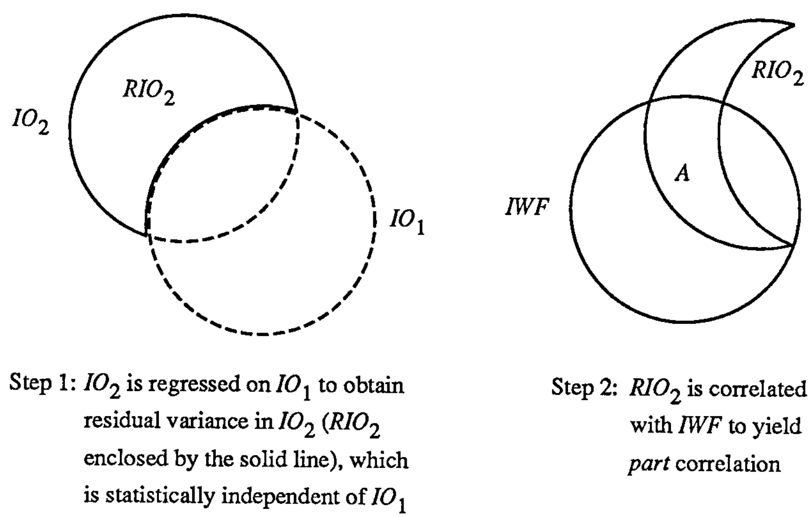
all associated to varying degrees with precollege intellectual orientation. Thus, the relationships between both institutional size and informal contact with faculty, on the one hand, and postfreshman intellectual orientation, on the other, are confounded by precollege intellectual orientation. (The reader is referred back to Figure A.1 for a visual representation of a confounded relationship.) The analytical issue, then, becomes one of estimating that part of the relationship between measures of college experience (for example, *IWF*) and college output (for example, IO_2) that is unconfounded by student background (for example, IO_1). This is represented by area *A* in Figure A.1.

Use of Residual Scores

One early approach to estimating the unconfounded relationship between college experience and college output has been advocated and refined by Astin, largely in connection with his influential input-environment-output model for assessing college impacts (Astin, 1963a, 1963b, 1965a, 1968b, 1968c, 1990; Astin & Panos, 1966, 1969). In this model, *input* refers to the traits or characteristics that students bring to college (precollege intellectual orientation in our example data), *environment* refers to the college environment or students' college experiences (institutional size and informal interaction with faculty in our example data), and *output* refers to the college outcome being explained or predicted (postfreshman intellectual orientation in the example data). The analytical procedure involves two steps. In the first step, student outputs are regressed on student inputs, and the regression equation produced is used to compute a predicted or expected output. This predicted output is then subtracted from actual or observed output to yield a residual output. Since predicted and residual scores are uncorrelated or statistically independent of each other, the residual score represents variation in output statistically purged or independent of the influence of input. Measures of the college experience are then correlated (step 2) with the residual output (RIO_2) to yield a "part" correlation representing the net or unique association of college experience and output. The square of this part correlation (r^2) represents the portion or percentage (when multiplied by 100) of the residual variance in output that is systematically shared with college experience but not with student input. This relationship is portrayed by area *A* in Figure A.2, where *IWF* represents the college experience variable and RIO_2 (enclosed by the solid line) represents the residual variance in IO_2, with variance shared with IO_1 removed statistically.

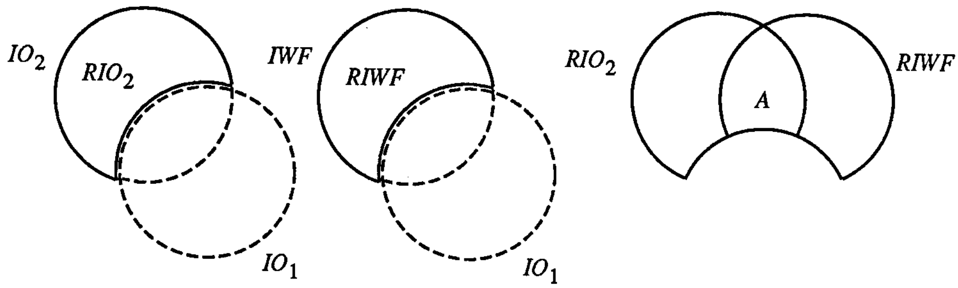
In our example, both the independent *(IWF)* and dependent (IO_2) variables are continuous in nature. It is also possible, however, to compute a part correlation between a categorical independent variable (for example, college exposure versus noncollege exposure) and the residual variance in a continuous dependent measure. This residual score/part correlation procedure with a categorical independent variable is typically referred to as the analysis of covariance (Cohen & Cohen, 1975).

Figure A.2. Use of Residualized Outcome Score (Part Correlation).



A related procedure using residual scores is the partial correlation. In the partial correlation both college experience *(IWF) and* output (IO_2) are residualized on input (IO_1). The two sets of residual scores represented geometrically by areas RIO_2 *and RIWF* would then be correlated. The square of the partial correlation would represent the percentage of residual variance in IO_2 (RIO_2) systematically associated with residual variance in *IWF (RIWF)*. In our example the squared partial correlation between *RIWF* and RIO_2, statistically controlling for IO_1, equals .10, or 10 percent, of the residual variance. A visual portrayal of the partial correlation is shown in Figure A.3. As the figure illustrates, the square of the partial correlation is area *A*, which corresponds to area *A* in the part correlation (see Figure A.2). Figures A.2 and A.3 differ only in that the total variance in *IWF* is correlated with the residual variance in IO_2 (RIO_2) in Figure A.2, and the residual variances in *IWF (RIWF)* and IO_2 (RIO_2) are correlated in Figure A.3. Comparing Figures A.2 and A.3, it is questionable that the partial correlation

Figure A.3. Partial Correlation.



Step 1: Both IO_2 and IWF are regressed on IO_1 to obtain residual variance in IO_2 (RIO_2) and IWF ($RIWF$), both of which are statistically independent of IO_1

Step 2: RIO_2 and RIWF are correlated to obtain *partial* correlation

provides any more stringent statistical control over confounding variables than does the part correlation (Cohen & Cohen, 1975). Indeed, both part and partial correlations will, in fact, have the same level of statistical significance.

As pointed out by Feldman (1971b) and Werts (1968), the residual score procedure, in both its part and partial correlation forms, statistically removes any joint association between input and the college experience. Consequently, the variance in output that could be *jointly* attributable to the influence of input and the college experience is automatically attributed to the variable partialed out (typically and in the present case, input). This means that the residual score procedure may overestimate the influence of input and provide a very conservative estimate of the influence of college experience. This may not always be undesirable, however. As suggested by Feldman (1971b), there may be situations in which the analyst has a sound theoretical rationale for attributing the shared portion of variance in student outcome to student input and intentionally does so. When this happens, value connotations attached to the words *overestimation* and *underestimation* are inappropriate (Feldman, 1971b, p. 135). Fuller discussions of the use of residual scores in computing and interpreting part and partial correlations are found in Astin (1970a), Creager (1969), Feldman (1971b), and Werts and Watley (1968a).

Clearly, the use of residual scores as a mechanism to gain some measure of contról over confounding input variables has been a major methodological advance in the study of college impact. There are, however, some problems with this approach. One is the reliability of residual scores, which may be a particular problem when one uses the two-step part correlation procedure (Kerlinger & Pedhazur, 1973). Perhaps more important, however, is the fact that the use of residual scores allows one to explain only the percentage of the residual variance in the criterion measure. This means that squared part or partial correlations based on residual scores do not represent the portion of total variance in college output uniquely associated with college experience. Rather, they represent association with that portion of the variance in college output that remains after covariation with input has been removed. Thus, the squared part and partial correlations based on residual scores will be larger in magnitude than they would be if the baseline were the total variance in the output criterion. How much larger sometimes depends on the correlation between input and output.

Multiple Correlation and Regression Procedures

Frequently, the analyst is interested in explaining or partitioning the total variance in some college outcome measures. The most useful set of approaches for accomplishing this is multiple correlation and regression analysis (for example, Cohen, 1968; Darlington, 1968; Draper & Smith, 1966; Kerlinger & Pedhazur, 1973; Pedhazur, 1975, 1982). This procedure allows one to estimate the percentage of total variance in a criterion measure systematically associated with a set of independent variables. It also allows for partitioning of the explained variance into different unique and common (or joint) components, depending upon the analyst's understanding of the conceptual model being estimated. Thus, in our present example the explained variance (R^2) in IO_2 is a function of IO_1, *SIZE*, and *IWF*. The actual R^2 achieved for our fictitious data is .4266, or 42.66 percent. This means that approximately 43 percent of the variance in IO_2 is systematically associated with our three independent variables and that approximately 57 percent ($1 - R^2$) is due to errors of measurement and other influences not specified in our regression model.

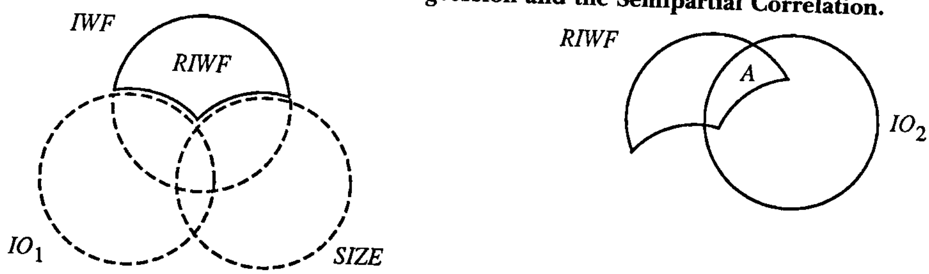
It is worth noting that the R^2 for our data (43 percent) is less than it would be if we had simply summed the squared simple correlations between each independent variable and IO_2 as shown in Table A.1; that is, $.60^2 + (-.15^2) + .33^2 = .49$. This difference is due to the intercorrelations, or multicolinearity, among the independent variables (also shown in Table A.1). Because of this multicolinearity, the independent variables are, at least in part, laying claim to some of the same variance in IO_2. The computation of R^2 is designed to take this redundant association into account. An excellent and readable discussion of the computations involved in obtaining R^2 is provided by Cohen and Cohen (1975).

Commonality Analysis

In addition to estimating the variance explained by our total set of three predictors, we are also interested in the proportion of the variance that can be attributed to each of the independent variables. One procedure that has received a reasonable amount of use is commonality analysis. The purpose of commonality analysis is to "identify proportions of variance in the dependent variable that may be attributed uniquely to each of the independent variables, and proportions of variance to be attributed to various combinations of independent variables" (Pedhazur, 1975, p. 252).

To obtain the unique variance estimate for each independent variable, a residualizing procedure analogous to that discussed above is carried out. In this procedure, however, each independent variable in the equation is residualized on all other independent variables to produce a measure that is purged of any association with other independent variables. These residualized independent variables are then correlated with the total variance in the criterion measure. This is portrayed in Figure A.4 for the relationship between *IWF* (residualized on IO_1 and *SIZE*) and IO_2. Figure A.4 shows the squared semipartial correlation (area *A*) between *IWF* (residualized on IO_1 and *SIZE*) and total variance in IO_2. Thus, it can be thought of as the unique or net relationship between *IWF* and IO_2, with IO_1 and *SIZE* controlled statistically. The net variance in IO_2 associated with IO_1 and with *SIZE* is computed in the same manner.

Figure A.4. Multiple Correlation/Regression and the Semipartial Correlation.



Step 1: Each independent variable, in this case *IWF*, is regressed on all other independent variables to obtain residual variance in *IWF* (*RIWF* enclosed by the solid line); *RIWF* is statistically independent of both IO_1 and *SIZE*

Step 2: Each residualized independent variable, in this case *RIWF*, is correlated with the *total* variance in the dependent variable (IO_2) to yield a semipartial correlation

Another way to conceptualize the unique or net criterion variance associated with a specific independent variable is to think of it as the proportion of the total variance that is explained above and beyond that associated with all other independent predictors. Consequently, if we are interested in the net or unique variance in IO_2 associated with *IWF*, we might compute two R^2s. The first would be the R^2 associated with the regression of IO_2 on IO_1 and *SIZE*, while the second would be the R^2 associated with the regression of IO_2 on IO_1, *SIZE*, and *IWF*. The difference between the R^2s of the two equations would represent the variance in IO_2 associated with *IWF*, *net* of IO_1 and *SIZE*. An analogous set of regression equations would be computed to estimate the unique variance in IO_2 associated with IO_1 and *SIZE*, respectively.

The joint or common variance among sets of predictor variables is essentially that which cannot be uniquely attributed to the individual independent variables. In our fictitious example the common variance is the difference between the R^2 in IO_2 associated with IO_1, *SIZE*, and *IWF* and the sum of the unique variance increments with each individual independent variable. Table A.2 shows the unique and common variance increments based on regression analysis of the correlation matrix in Table A.1. As Table A.2 shows, of the 42.66 percent of the total variance in IO_2 explained by $IO_1 1$, *SIZE*, and *IWF*, 5.35 percent is the common variance or that which could not be attributed uniquely to any of the three independent variables.

Table A.2. Unique and Common Variance Increments.

Source	*Increment*
IO_1	31.47%
SIZE	0.19%
IWF	5.65%
Common variance increment	5.35%
Total variance explained (R^2)	42.66%

It is also possible to calculate the common or joint variance estimates in a dependent measure that can be attributed to any particular subset of independent variables that may be of interest. Reasonably straightforward rules for writing commonality formulas that use various combinations of simple regression equations are provided by Kerlinger and Pedhazur (1973), Mood (1969, 1971), and Wisler (1969).

Commonality analysis has a great deal of intuitive appeal and at first glance seems a perfectly logical and straightforward method for determining the relative importance of a variable in explaining criterion variance. Needless to say, however, there are problems with such a conclusion. A major limitation of commonality analysis is that the unique variance estimates are highly influenced by the degree of intercorrelation or multicolinearity among the independent variables (Creager, 1969). Although not particularly acute in our fictitious example, it is often the case that student input characteristics are substantially correlated with measures of both college characteristics and the college experience. (This is particularly true when institutions are the unit of analysis.) In the situation where the independent variables are themselves highly intercorrelated with each other, it is not uncommon to get large increments of joint variance and relatively trivial increments of unique variance. This makes it quite difficult to talk about the relative importance of a specific variable. Because of the way in which the unique contribution of a variable is affected by its intercorrelations with other independent variables, there is some question as to its appropriateness for assessing the substantive importance of individual variables in regression analyses (Astin, 1990; Pedhazur, 1975; Tukey, 1954).

As suggested by Pedhazur (1975), the critics of variance partitioning have questioned both its validity and its usefulness in terms of policy decisions. As an alternative, they have advocated a regression coefficient approach, which has as its purpose the estimation of the "effects" of each of the independent variables on the criterion measure (for example, Blalock, 1968; Linn, Werts, & Tucker, 1971; Werts, 1970; Werts & Watley, 1968a). In our fictitious example, the regression equation would be as follows:

$$IO_2 = a + b_1\,(IO_1) + b_2(SIZE) + b_3(IWF) + \text{error},$$

where a = the intercept or constant; b_1, b_2, and b_3 = the regression coefficients associated with the independent variables or predictors; and error = causes of IO_2 not specified in the equation plus errors of measurement.

In the multiple regression case, the regression coefficients or weights come in two forms. In the equation shown above, the b weights are metric (or unstandardized) partial regression coefficients. Thus a b weight indicates the expected change in IO_2 for a unit increase in the independent variable with which it is associated, while the influence of all other independent variables in the equation is statistically held constant. In their standardized form, the beta regression weights indicate the expected standard deviation change in the dependent measure (IO_2) associated with one standard deviation increase in an independent variable, while all other variables are held constant (Pedhazur, 1975). As such, b weights are in raw-score terms and to some extent reflect the relative scale or metric of the independent and dependent variables. Beta weights, on the other hand, put all variables on the same scale and thus provide a scale-free estimate of a variable's effect on the dependent measure. In our fictitious data, the respective metric (b) and standardized (B) coefficients are as follows:

	Metric (b)	*Standardized* (B)
IO_1	.590	.564
SIZE	−.0001	−.046
IWF	.727	.246

Which of the two regression coefficients (b or B) should bear the interpretative burden in regression analysis has been the topic of considerable debate (for example, Blalock, 1964, 1968; Bowles & Levin, 1968; Cain & Watts, 1970; Hanushek & Kain, 1972; Schoenberg, 1972; Werts & Watley, 1968a). A principal problem of beta (B) weights is that they are not particularly stable across different samples, whereas *b* weights are. Thus, in comparing the relative effect of a variable *between* two different samples, the metric *b* weight is preferable (Pedhazur, 1982). Similarly, because the beta weight is essentially scale-free, there is some doubt about its usefulness in terms of policy. For example, within the same sample, relative beta weights may indicate that years of formal education are more important in explaining annual salary (say, in dollars) than is intelligence. One cannot, however, use the beta weight to estimate the average income benefits of one additional year of education, net of other factors. To do this, one needs to rely on the *b* weight. Because the *b* weight represents the partial influence of an independent variable on a dependent variable in terms of the actual units in which they are measured, it provides a direct estimate of the average dollar increase in income associated with each additional year of formal education (for example, Linn, Werts, & Tucker, 1971; Pedhazur, 1975; Smith, 1972).

A counterargument can be made, however, concerning the relative usefulness of *b* and beta weights in assessing the relative contribution or effects of independent variables *within* a sample. In this case, the metric regression weight is limited because of the very fact that its magnitude is a function of the scale units employed in measuring independent and dependent variables. Consequently, the relative magnitudes of *b* weights for several independent variables may have little correspondence to the relative importance of those variables in terms of effects on the dependent measure. In the schooling-intelligence-income illustration, for example, schooling may be measured on a 1–17 scale (where 16 = a bachelor's degree and 17 = some graduate study; or 0–1, where 1 = college graduate and 0 = nongraduate), while intelligence may be measured on a much more variable scale, say, 90–150.

When one is primarily interested in the relative effects of different independent variables *within the same sample,* it may be more appropriate to allow the beta or standardized regression weights to bear the interpretative burden. On the other hand, if one is interested in comparing the relative effects of the same variable *between or among different samples* (for example, between men and women), the metric regression weight is a more useful index (Blalock, 1967; Kerlinger & Pedhazur, 1973; Pedhazur, 1975, 1982; Smith, 1972; Wright, 1960).

Regression Coefficients and Causality

There is a tendency in much literature on the influence of college to speak of regression coefficients as estimates of the net effect of the independent variable on a dependent measure. Given the above definitions of metric and standardized regression coefficients, this seems a natural interpretative conclusion. In an important discussion of the meaning of regression coefficients, however, Pedhazur (1975, 1982) makes an important distinction between the causal meaning of regression coefficients derived from experimental data and those yielded by correlational data.

The easiest way to illustrate this distinction is through an admittedly contrived example. Suppose we randomly provide half the entering freshmen in a particular college with a dictionary-thesaurus combination and withhold it from the remaining half. At the end of the freshman year we give the entire class a test of vocabulary and find that those who received the dictionary thesaurus scored significantly and substantially higher than the nonrecipients. The unstandardized regression equation would tell us the average advantage in vocabulary test achievement accruing to those freshmen provided with the dictionary thesaurus (group coded 1) versus those not provided with it (group coded 0). Given this randomized, true experiment, we could estimate (by means of the *b* weight) the typical improvement in vocabulary achievement we might get by routinely providing all incoming freshmen with a dictionary thesaurus.

Conversely, suppose in a correlational, panel study we find that net of precollege level of vocabulary achievement, having a dictionary thesaurus has a net positive association with vocabulary achievement at the end of the freshman year. In this situation we have not been able to manipulate and control the conditions under which the relationship between having a dictionary thesaurus and outcome vocabulary achievement is observed. Consequently, the regression coefficient allows us *only* to estimate the average difference in vocabulary achievement between freshmen who own a dictionary thesaurus and those who do not, net of precollege vocabulary achievement. We cannot tell from the regression coefficient whether purposefully providing freshmen with a dictionary thesaurus would produce the same effect (Pedhazur, 1975).

In short, we cannot interpret regression coefficients from naturally occurring correlational data as though the variables they are associated with had been purposefully manipulated under experimental conditions. It would be misleading, therefore, to interpret them as the change in the dependent variable that we can *expect* from a purposeful unit increase in the independent variable, net of other independent variables. Regression coefficients from correlational data can be quite useful in identifying *possibly* causal associations among variables. In the vast majority of investigations on the influence of college, however, the regression coefficients are, in and of themselves, insufficient evidence for causality.

Hierarchical and Stepwise Regression Analysis

In the approach to regression or commonality analysis discussed above, all independent variables in the model are entered into the equation at the same time. It is worth briefly mentioning two other approaches to regression analysis that are concerned with the order in which independent variables enter a regression equation. The first is termed hierarchical regression, and it is often used when the researcher can posit an explicit causal or temporal ordering or hierarchy to the independent variables (for example, Chapman & Pascarella, 1983; Pascarella & Terenzini, 1980a; Terenzini & Pascarella, 1978).

For example, suppose in our fictitious data we are interested only in whether measures of the college environment and experience (for example, institutional size and informal contact with faculty) account for a substantive increase in the explained variance in freshman-year intellectual orientation over and above the individual's level of precollege intellectual orientation. Consistent with our question, we would solve the regression analyses in two steps. In the first step, IO_2 would be regressed on IO_1, and in the second step, *SIZE* and *IWF* would be added to the equation. The increase in R^2 from step 1 to step 2 would represent the unique contribution of college environment and experience variables (*SIZE* and *IWF*) over and above that due to input (IO_1). (In the actual data, the R^2 in IO_2 increases from .360 with only IO_1 in the equation to .426 with *SIZE* and *IWF* added, or an R^2 increase due to *SIZE* and *IWF* of .066 or 6.6 percent.) Hierarchical analysis makes no attempt to find the joint or common variance due to input (IO_1) and college environment and experience (*SIZE* and *IWF*). Rather, it attributes any joint influence to input alone. As such, hierarchical approaches provide a conservative or lower-bounds estimate of the output variance associated with college environment or experience. This, however, may accurately reflect the analyst's conceptualization of the process.

Although hierarchical analysis enters variables or sets of variables in steps, it does so in accordance with a conceptual or causal hierarchy specified by the researcher. Thus, it must be distinguished from another type of analysis, termed stepwise regression, with which it is sometimes confused. In stepwise analyses, variables are entered (or removed) in steps, but the criteria and order of entry (or removal) are empirically rather than conceptually determined. In one type of stepwise analysis, for example, the computer searches first for the independent variable with the largest simple or zero-order correlation with the dependent measure and enters it into the regression equation in step 1. In step 2, the computer searches the *remaining* independent variables and selects the one with the largest partial correlation with the dependent variable (net of the variable already in the equation) and adds it to the equation. This step-by-step, empirically driven procedure continues until a specified criterion for entry is no longer met (for example, the R^2 increase is no longer statistically significant or is less than 1 percent, and so on).

The basic purpose of stepwise analysis is to develop a parsimonious accounting of variance in the dependent variable (that is, an optimal accounting with the fewest and most important predictors). This has a lot of intuitive appeal, but stepwise analysis has some serious limitations. The most important of these is that it capitalizes on chance covariation in a sample. The easiest way to show this is through an example. Suppose in sample 1 we have ten independent variables, but the two variables with the highest simple correlations with the dependent measure are *A* (r = .31) and *B* (r = .30). In sample 2, with the same set of independent variables, the two respective highest correlations are *A* = .30 and *B* = .31 (just the reverse). In a stepwise analysis conducted with sample 1, variable *A* would enter the equation first and, if *A* and *B* are highly correlated, *B* might never enter the equation. In sample 2, variable *B* would enter first, and if *B* and *A* are correlated sufficiently, *A* might never be selected for inclusion. Moreover, in sample 1 the variables entering the equation would be strongly determined not only by their correlation with the dependent variable but also by how much they covary with *A*. In sample 2, however, an important factor determining the final equation would be correlations between the remaining independent variables and variable *B*. If these respective correlation patterns differ, the result could be quite different equations in sample 1 and sample 2 even when the same independent variables are being considered.

What this comes down to is that regression equations determined by stepwise analysis can be unstable across independent samples. Indeed, as a rule of thumb, it is probably wise to trust stepwise-determined equations only when the results have been validated on another independent sample (Kerlinger & Pedhazur, 1973).

Causal Modeling

Regression analysis provides the same level of statistical control as the use of residual scores and partial correlations. In addition, it permits partitioning of the total variance in the dependent measure and provides (through regression coefficients) ways of estimating the net magnitude of the influence of each independent variable on the dependent measure. Due in large part to these and related advantages, regression analysis has generally replaced the use of residual scores and partial correlations in studies concerned with the influence of college.

Despite its utility as a general data analytical system, however, regression analysis in the forms discussed above is largely predictive rather than explanatory in nature. For example, when we regress IO_2 on IO_1, *SIZE*, and *IWF* in our fictitious data set, the resultant regression equation allows us to determine the unique and joint variance increments in IO_2 associated with the three predictors. It also allows us to estimate the net change in IO_2 associated with unit changes in each of the three predictors. It does not, however, provide much information in terms of explaining the interactive process through which student precollege traits (for example, IO_1), institutional characteristics (for example, *SIZE*), and individual collegiate experiences (for example, *IWF*) influence one another as well as freshman-year outcomes (for example, IO_2). Moreover, simple regression analysis does not really help us understand the various mechanisms through which the joint or common influences among independent variables may have an effect on the dependent measure.

Causal modeling is a use of regression analysis that focuses on explanation rather than prediction and provides an efficient method for determining the indirect as well as the direct influences of each independent variable in a theoretically guided causal system (Anderson & Evans, 1974; Duncan, 1966, 1975; Feldman, 1971a; Heise, 1969; Werts & Linn, 1970; Wolfle, 1977, 1980a; Wright, 1934, 1954). As argued by Wright (1934) and discussed by Wolfle (1985a), the purpose of causal modeling is not to accomplish the impossible (that is, to attribute experimental causality from correlational data). Rather, its purpose is to determine the extent to which an *a priori* system of hypothesized causal effects is supported by actual data. Thus, causal relationships exist in models not because they are "proven" by regression coefficients but because theories posit them. From this perspective, if regression coefficients tend to support the presence of a hypothesized causal relationship, this suggests only the *possibility* that the observed relationship may be causal. One does not confirm a causal relationship with regression coefficients in causal modeling; one only "fails to disconfirm it" (for example, Cliff, 1983; Kenny, 1979; Wolfle, 1985a).

As suggested in a comprehensive and readable discussion of causal modeling by Wolfle (1985a), the use of causal models in research on the influence of college may have two important advantages. First, causal modeling requires that the researcher give considerable thought to the theoretical structure of the problem. This means that he or she must not only specify the relevant independent and dependent variables but must also be explicit about the presumed causal ordering and patterns of cause and effect in the model. Often this is accomplished by drawing a path diagram of the causal model in which causal arrows reflect relationships among variables suggested by theory or relevant literature. The explicitness of causal modeling can be important in preventing the misinterpretation of results. As Wolfle (1985a, p. 383) puts it, the "researcher may, of course, be wrong but he won't be misunderstood."

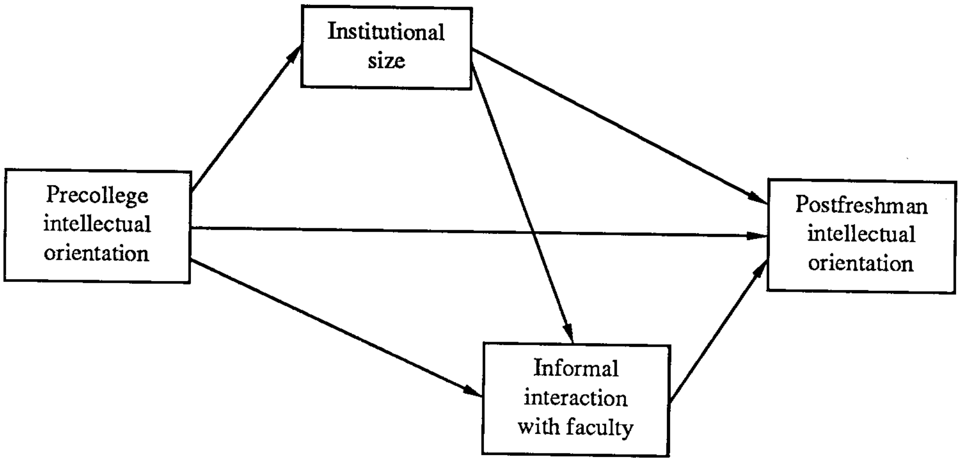
The second advantage of causal modeling is that it moves beyond typical regression analysis and allows the researcher to investigate not only the direct unmediated causal effects of each independent variable but also their indirect effects through intervening variables in the model (Finney, 1972; Wolfle, 1985a). It is in the investigation of indirect effects that causal modeling provides substantively more information than do typical regression analyses. The estimation of direct effects in a causal model can be achieved through the simple regression of the dependent variable on all independent variables in the model. Thus, direct effects in causal models are the same as standardized regression weights in regression analysis, and both are analogous to the unique variance estimates in commonality analysis. The indirect effects in causal modeling, however, can be thought of as a way to further understand and disaggregate the joint or common variance among independent variables (Alwin & Hauser, 1975; Feldman, 1971a).

Decomposition of Effects in Causal Modeling. There are three types of effects in causal modeling. (Note that the terms *path analysis* and *structural equations modeling* are often used interchangeably with *causal modeling.*) These types are direct effects, indirect effects, and total effects. The direct effects have been defined above, the indirect effects are the sum of the products of direct effects through intervening variables, and the total effects are the sum of the direct and indirect effects. The difference between the simple or zero-order correlation between an independent and a dependent variable and the total effect of that independent variable can be considered the spurious or noncausal part of the relationship.

A simple way to understand indirect effects is to visualize three billiard balls: *A, B,* and *C.* If ball *A* strikes ball *B,* which in turn strikes ball *C,* then ball *A* can be thought of as having an indirect effect on ball *C* through ball *B.* Thus ball *A* has an effect on ball *C* even though it may never strike it directly. Similarly, college grades may not directly influence income. Nevertheless, they clearly have a strong positive influence on degree attainment, which in turn is a key determinant of income. Consequently, grades in college may have an important indirect effect on income through educational attainment.

The easiest way to demonstrate the decomposition of effects in a causal model is through the use of an example. Again, we will use the fictitious correlation matrix found in Table A.1. The difference, however, is that this time we will posit a hypothetical causal structure to the variables. Let's suppose that on the basis of some theory, we hypothesize that the enrollment of the institution attended *(SIZE)* is a function of precollege level of intellectual orientation, with students initially high in intellectual orientation tending to select and attend smaller schools. In turn, informal contact with faculty *(IWF)* is hypothesized as being a function of both precollege intellectual orientation and institutional size. The former is hypothesized as positively influencing informal contact with faculty, while institutional size is hypothesized to inhibit faculty contact. Finally, end-of-freshman-year intellectual orientation (IO_2) is posited as being positively influenced by IO_1 and *IWF* and negatively influenced by *SIZE.* A visual portrayal of the model is shown in Figure A.5. In the model, IO_1 is considered an exogenous variable because it is determined by effects outside the model. *SIZE, IWF,* and IO_2 are considered endogenous variables and are determined by exogenous variables and all other causally antecedent endogenous variables in the model.

Figure A.5. Proposed Causal Model.

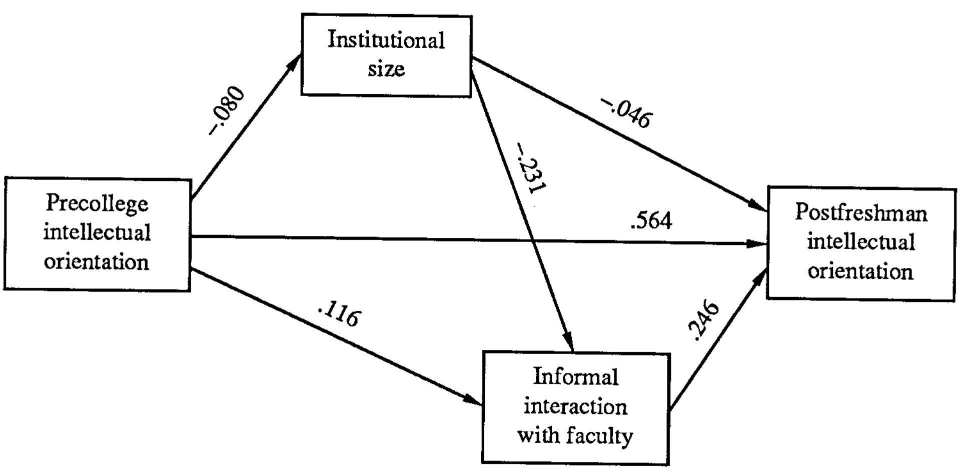


In order to estimate the direct and indirect effects specified by the model, it is necessary to solve a series of regression equations that define the structure of effects in the model. These "structural equations" for the model shown in Figure A.5 are as follows:

$$SIZE = a + B_1(IO_1) + \text{error}$$
$$IWF = a + B_1(IO_1) + B_2(SIZE) + \text{error}$$
$$IO_2 = a + B_1(IO_1) + B_2(SIZE) + B_3(IWF) + \text{error},$$

where B equals the standardized regression or path coefficient and a is a constant. Thus, the structural equations provide all the direct effects, which in turn can be used to compute indirect effects. Figure A.6 shows the ap-

Figure A.6. Path Coefficients.



propriate direct effects (beta weights) for each causal path computed for our fictitious correlation matrix. The decomposition of causal effects on IO_2 is summarized in Table A.3. As this table shows, the direct or unmediated effects of IO_1 (.564) and *IWF* (.246) on IO_2 were positive and substantial; both were also statistically significant. The direct effect of *SIZE* (-.046), however, was small and nonsignificant.

Had we ended our analysis here, essentially the results of a simple multiple regression, we would have concluded that both precollege intellectual orientation and informal contact with faculty may causally influence freshman-year intellectual orientation but that the size of the institution attended is essentially unimportant. This would have been misleading, however, since as we can see from Figure A.6, *SIZE* has a substantial (-.231) negative effect on *IWF*, which in turn positively influences IO_2 (.246). Based on our computational definition of indirect effects (the sum of the products of direct effects through intervening variables), the indirect effect of *SIZE* on IO_2 through $IWF = (-.231) \times (.246) = -.057$. Though modest, this indirect effect is larger than the direct effect and is statistically significant, while the latter is not. (The statistical significance of indirect effects can now be computed quite simply via a computer program developed by Wolfle and Ethington, 1985, based on the work of Sobel, 1982.)

This example demonstrates an important analytical advantage of causal modeling over simple regression analysis. By permitting one to estimate indirect (mediated) as well as the direct (unmediated) effects, it provides a more complete estimate of a variable's total influence on the criterion. In this sense it addresses and helps illuminate the nature of a variable's influence through its joint covariation with other independent variables (that is, the joint or common variance from a commonality analysis). In doing so, it may alter one's conclusions about the importance of a variable's impact versus the estimate one would get from an examination of direct effects only. In the present case, the total effect of *SIZE* on IO_2 —that is, the sum of the direct (−.046) and indirect (−.057) effects—is −.103, which is also statistically significant. This suggests that in our hypothetical example, the enrollment of the institution attended has a modest negative influence on the development of intellectual orientation, primarily by inhibiting student informal contact with faculty. Our conclusion based on direct effects only

Table A.3. Direct, Indirect, and Total Effects.

Variable	*Direct Effect*	*Indirect Effect*	*Total Effect*
IO_1	.564[a]	.036	.600[a]
SIZE	−.046	−.057[a]	−.103[a]
IWF	.246[a]		.246[a]

[a] $p < .01$.

would have essentially dismissed any influence of *SIZE* on IO_2 as trivial and statistically nonsignificant.

It is worth noting that there are different ways of computing indirect and total effects in a causal model. Indirect effects can be computed as shown above. In complex models, however, the summing of the products of direct effects through the various paths leading to the dependent measure can be tedious and fraught with computational errors. A simpler way is to compute direct effects and total effects and then subtract the former from the latter to get indirect effects. Direct effects, of course, can be obtained from the simple regression of the dependent measure on all relevant independent measures. Total effects and their statistical significance can be obtained by using "reduced form" regression equations (Alwin & Hauser, 1975). These are simply the regression of the dependent measure on the variable of interest and all causally concurrent or antecedent variables in the model. Intervening variables between the variable of interest and the dependent measure are not in the equation—thus the term *reduced form.* For example, the total effect of IO_1 on IO_2 is simply the regression of IO_2 on IO_1. The independent variable here is exogenous, so it has no causal antecedents defined by the model. The total effect of *SIZE* is obtained from the regression of IO_2 on IO_1 and *SIZE*. Finally, since *IWF* has no indirect effect on IO_2 through intervening variables in the model, its direct effect is also its total effect.

Problems in the Use of Causal Modeling. While causal models have great potential for increasing the information gained about the process of collegiate influence on student development, a number of problems are inherent in their use. The first of these focuses on the issue of adequate model specification. This means that in constructing causal models the researcher must give considerable thought both to including all important relevant variables in the model and to specifying their appropriate causal ordering. If important causal influences are excluded, the result is often seriously biased or inflated path coefficients. For example, with IO_1 and *SIZE* also in the equation, the direct effect of *IWF* on IO_2 is .246. However, if IO_1 is left out, the direct effect of *IWF* on IO_2 jumps spuriously to .312. On the other hand, if the model specifies an inappropriate or implausible causal ordering in the variables, the indirect effects can be misleading and perhaps even meaningless.

What determines an adequately specified model, of course, is the soundness of its theoretical structure. Indeed, as noted by Duncan (1975, p. 149), quoted in Wolfle (1985a), "The study of structural equation [causal] models can be divided into two parts: the easy part and the hard part." The easy part is the solution of structural equations and the computation of direct, indirect, and total effects. The hard part is the construction of causal models that reflect sound social theory.

In addition to adequate model specification, a second assumption of causal modeling, and indeed of ordinary least-squares multiple regression generally, is recursiveness. *Recursiveness* refers to the assumption of unidirectional causal influence (for example, A influences B but not vice versa) in the model and the absence of causal feedback loops. Thus, in the example data, informal contact with faculty is presumed to influence freshman-year intellectual orientation, but the reverse is not hypothesized. The only way to ensure that causal feedback loops (or correlated errors) are not present is to collect longitudinal data in a manner that reflects the causal ordering of variables in the model. For example, in our fictitious study, measures of IO_1 were collected prior to measures of IWF, which in turn preceded the collection of data relevant to IO_2. This means that while IWF may causally influence IO_2, the reverse is not possible (for example, Kenny, 1979). If, however, independent measures such as IWF and dependent measures such as IO_2 had been collected simultaneously on the same instrument (a practice quite common in educational research), the direction of causal effects could be ambiguous. Does informal interaction with faculty influence intellectual orientation, or do increases in intellectual orientation lead students to seek out more frequent nonclassroom interaction with faculty?

Consistent with what we know about the developmental nature of student growth and maturation during college (for example, Chickering, 1969; Feldman, 1972; Heath, 1968), causal models assume longitudinality in the data being analyzed. In the absence of carefully collected *longitudinal* data, it is extremely hazardous to assume the presence of unidirectional causal effects. Moreover, when there is simultaneous assessment of the dependent variable and independent causal variables on the same instrument (that is, cross-sectional data), the frequent result is a correlation between those independent causes and errors of measurement on the dependent variable. This violates one of the assumptions of any ordinary least-squares regression analysis and can lead to biased regression coefficients (Pedhazur, 1982).

This problem has led to the development of nonrecursive causal models in which the researcher actually hypothesizes the presence of causal loops or two-way causality (Anderson, 1978). Estimating nonrecursive causal models requires the use of a procedure termed two-stage least-squares regression analysis. One of the assumptions of two-stage least squares is the presence of "instrumental" exogenous variables. By definition, an instrumental exogenous variable is one that causally influences one of the dependent (endogenous) variables in the causal loop but has zero effect on the other (Anderson, 1978; Wolfle, 1985a). Once these instrumental exogenous variables are identified and the causal model is specified, existing two-stage least-squares regression programs, such as those available in the Statistical Analysis System (SAS), can be used to estimate the direction and strength of causal influences. Unfortunately, given the considerable correlations among student precollege traits, institutional characteristics, and measures of the collegiate experience in research on college influence, it is extremely difficult to find exogenous variables that meet the assumption of also being instrumental variables. Wolfle (1985a, p. 395) has suggested that this problem may make "most nonrecursive models implausible for social science applications." Nevertheless, there are a few uses of nonrecursive models and two-stage least-squares regression in research on the college student (for example, Bean & Bradley, 1986; Bean & Kuh, 1984; Iverson, Pascarella, & Terenzini, 1984).

A third important issue in the use of causal models is measurement error. One of the assumptions of any regression analysis is that both the independent and dependent variables are measured without error; however, this is almost never the case. In the presence of measurement error, regression or path coefficients will be biased, and it is extremely difficult to determine both the magnitude and the direction of the biases (Maruyama & Walberg, 1982; Stage, 1989, 1990). Recently, however, there have been a number of advances with maximum likelihood estimation procedures such as linear structural relations (LISREL) (Joreskog, 1973; Joreskog & Sorbom, 1979, 1983). These procedures permit correction for differential measurement errors and thus yield less biased regression coefficients than do ordinary least-squares regression approaches. As such, they represent an important refinement in the application of causal modeling to questions of the influence of college on student development.

The Unit of Analysis

An important question in investigations of the influence of college on student development is the appropriate unit of analysis. This is most apparent when one is analyzing multi-institutional samples where data are collected at the individual level but where it is also possible to obtain average scores at the institutional or other level of aggregation (for example, the Cooperative Institutional Research Program data, the National Longitudinal Study of the High School Class of 1972, High School and Beyond). It is also possible, however, that one needs to consider the appropriate unit of analysis even when the data come from single-institution samples. What, for example, are the effects of classroom climate or residential unit composition on student learning (for example, Pascarella & Terenzini, 1982; D. Smith, 1977; Terenzini & Pascarella, 1984)?

The unit of analysis issue has been a complex and somewhat controversial one in research on the influence of college. It is often the case that scholars interested in essentially the same question have in various studies used institutions, departments, or individuals as the unit of analysis. Consider, for example, studies of the influence of different college characteristics on student learning (for example, Astin, 1968c; Ayres, 1983; Ayres & Bennett, 1983; Centra & Rock, 1971; Hartnett & Centra, 1977). Variation in the unit of analysis has perhaps contributed to the lack of consistent findings in several areas of inquiry (for example, Pascarella, 1985a). As suggested in a sophisticated and cogent discussion by Burstein (1980a), the issue is not so much that one unit of aggregation is more appropriate than another. Rather, the issue needs to be understood in light of the fact that different units of aggregation or analysis are asking different questions of the data. When the institution is the unit of analysis, for example, one is essentially asking what the average influence of certain college characteristics (student body selectivity, average faculty salary, and so on) is on average student development. Thus, one is primarily concerned with average effects among or between institutions. When individuals are the unit of analysis, however, the question is typically whether differences in individual students' collegiate experiences (for instance, academic major, extracurricular involvement, interaction with faculty) lead to differences in specified outcomes. Here the focus is on the effects of different experiences or exposures among or between individual students, even if the data are multi-institutional in form.

By focusing on one question, both institutional and individual levels of aggregation tend to ignore other questions. Aggregating at the level of the institution tends to mask possibly substantial variations between individual students' experiences within the same institution (Cronbach, 1976). Assuming, for example, that an aggregate or global measure of the college environment accurately portrays a homogeneous stimulus experienced by all students in the institution ignores substantial evidence of influential subenvironments in an institution, subenvironments that are more proximal to the student's daily experience (for example, Baird, 1974; Berdie, 1967; Lacy, 1978; Newcomb & Wilson, 1966; Pascarella, 1976; Phelan, 1979; Weidman, 1979). Conversely, using individuals as the unit of analysis tends to ignore the dependencies (or correlations) of individual subject experiences within institutions; that is, the shared educational experience among individual students within the same college leads to the nonindependence of individual behaviors within the college (Burstein, 1980a). Thus, for example, institutional enrollment (size) may facilitate certain types of student-faculty relationships in a small liberal arts college that are quite different from the nature of the student-faculty relationships typically found in large research universities. These types of relationships may differentiate small and large institutions even when individual differences in student characteristics are taken into account. Moreover, as suggested by Burstein (1980a), standard statistical estimation techniques such as ordinary least-squares regression analysis can yield flawed or biased estimates in the presence of within-group dependencies.

Because of the dilemmas inherent in choosing one level of aggregation or unit of analysis over another, a number of scholars have suggested the appropriateness of using multilevels of analysis guided by appropriate theory (for example, Astin, 1970b; Burstein, 1980a, 1980b; Cooley, Bond, & Mao, 1981; Cronbach, 1976; Cronbach & Webb, 1975; Terenzini & Pascarella, 1984; Rogosa, 1978). In such analyses, both between-student and between-aggregation effects could be estimated when one has multi-institutional (or even multimajor, multiclassroom, or multiresidential arrangement) data. (The appropriate level of aggregation, of course, depends on the substantive question being asked.) Routine use of a multilevel approach such as this might be one way to permit a more valid and informative comparison of results across studies. It would also permit one to compare differences in the aggregate effects of college (or some other unit of aggregation) with the effects of individual student characteristics and experiences. As suggested by Burstein (1980a), variables can have different meanings at different levels of analysis. Studies that choose colleges as the unit of analysis are asking different questions than studies that use the individual as the unit of analysis; consequently, we should expect different results.

Contextual Analysis

One way of combining aggregate and individual levels of analysis simultaneously is through a procedure known as contextual analysis. Contextual analysis is essentially the study of the influence of group- or aggregate-level variables on individual-level outcomes (Erbring & Young, 1980; Firebaugh, 1978, 1980; Lazarsfeld & Menzel, 1961). In this procedure the individual is the true unit of analysis, but instead of focusing only on the developmental effects of individual college experiences, one also attempts to estimate the effect of being a member of a particular group or aggregation (for example, college academic major, residential unit, classroom).

In its simplest form, contextual analysis can be defined by the following regression equation:

$$Y_{ij} = a + b_1X_{ij} + b_2\overline{X}_j + \text{error},$$

where Y_{ij} might represent the academic achievement (for instance, Graduate Record Examination Scores) of the i^{th} student in college j, X_{ij} might be a measure of academic aptitude for the same student, and $\bar{X}_j$ would be the average (mean) value of student academic aptitude in college j. In short, X_{ij} might be thought of as a measure of student input or background, while $\bar{X}_j$ could be considered an estimate of college context or environment. The error or random disturbance term represents errors of measurement plus all causes of Y_{ij} (achievement) unspecified by the equation, such as student motivation and efficiency of study habits (Hanushek, Jackson, & Kain, 1974). The coefficients a (constant) and b_1 and b_2 (regression coefficients) can be estimated by ordinary least-squares regression procedures. A contextual or environmental effect is said to occur in this equation if the aggregate measure of student body aptitude has a significant regression coefficient with individual GRE achievement net of individual aptitude. If the coefficient for the contextual effect is positive, it would suggest that attending a college with a student body composed of "bright" students tends positively to influence a student's standardized academic achievement above and beyond his or her own academic aptitude.

One might posit the causal mechanism underlying the above example as due to the tendency for college faculty to gear the cognitive and conceptual level of instruction to the academic capacities of the students being taught or to the generally "higher" intellectual level of student discourse inside or outside the classroom. Hypothetically, then, students in more selective colleges might benefit from instruction (or an overall environment) geared to higher-level cognitive processes such as analysis, synthesis, and evaluation (Bloom, 1956), the results of which are manifest in higher GRE scores. In positing such a causal mechanism, however, we are again confronted by the disconcerting likelihood that selection (input) and aggregation (environmental) effects are substantially correlated. As such, it is extremely difficult, if not impossible, to accurately estimate and separate the effects of the latter from those of the former (Cronbach, Rogosa, Floden, & Price, 1977; Werts & Linn, 1971). In the above specification of the model, the unique effects of context or environment (as indicated by b_2) are likely to be quite conservative. Thus, a significant regression coefficient for average student body aptitude is reasonably convincing evidence of a unique contextual or environmental influence (Burstein, 1980a).

Frog-Pond Effects

Another approach to the combining of individual and aggregate level data is the "frog-pond" or relative deprivation effect as suggested by the work of Davis (1966), Alexander and Eckland (1975b), and Bassis (1977). This approach suggests that in order to understand individual behavior, one needs to be cognizant not only of individual attributes but also of how individual attributes position one in relationship to an important reference or peer group. In the above example of aptitude and achievement, the regression equation might be specified as follows:

$$Y_{ij} = a + b_1X_{ij} + b_2(\overline{X}_j - X_{ij}) + \text{error}$$

In this specification, hypothetical GRE achievement for an individual student (Y_{ij}) is posited as a function of individual academic aptitude (X_{ij}) and the difference between individual aptitude and the average college aptitude ($\bar{X}_j - X_{ij}$). A significant regression coefficient would indicate that a student's academic ability relative to the student average at the college attended has an influence on GRE achievement above and beyond individual aptitude alone. The sign of the regression coefficient would indicate whether the effect is generally beneficial to students below (+) or above (−) the college average.

As demonstrated by Burstein (1980a, 1980b), a regression equation including individual, contextual, and frog-pond effects is not estimable by standard means because the variables representing the three effects have a linear dependency. (The coefficients in an equation with any combination of two of the three effects represented, however, can be estimated.) Burstein (1980a) has suggested a way to deal with this problem. Specifically, he suggests that the investigator obtain more direct measures of the contextual or frog-pond effects. This means giving considerable thought to the specific and underlying causal mechanisms at work. For example, research conducted by Terenzini and Pascarella (1984) found that net of individual levels of institutional commitment, freshman-year persistence was independently and positively influenced by the average level of institutional commitment in the student's residence unit (contextual).

It is also possible that the student's level of institutional commitment relative to that of his or her residential unit peers (frog pond) would add significantly to an understanding of individual persistence or withdrawal behavior (an influence not estimated by Terenzini & Pascarella, 1982). That this effect operates through the influence of social involvement or integration is suggested by the theoretical work of Tinto (1975, 1982, 1987). Thus, instead of entering an unestimable frog-pond term operationalizing the student's standing relative to the average institutional commitment of the residence unit, one could substitute relative standing on level of social involvement. The equation then might be specified as follows:

$$P_{ij} = a + b_1\,(IC_{ij}) + b_2\,(\overline{\mathrm{IC}}_j) + b_3\,(\overline{\mathrm{SI}}_j - SI_{ij}),$$

where

P_{ij} = an individual student's persistence or withdrawal behavior,
IC_{ij} = an individual student's level of institutional commitment,
$\overline{IC}_j$ = average level of institutional commitment in a particular student's residence unit,
$(\overline{SI}_j - SI_{ij})$ = an individual student's level of social integration relative to the average in his or her residence unit.

Burstein's (1980a) argument for focusing on direct measures of aggregate and/or frog-pond effects underscores a major conceptual problem in multilevel analysis. This problem, which has been forcefully articulated by analysts such as Hauser (1970, 1974) and Firebaugh (1978), is that contextual or frog-pond effects estimated atheoretically are often mechanistic and distally related to the underlying social-psychological processes they were designed to represent (Burstein, 1980a). For example, contextual or frog-pond effects estimated at the institutional level may have little relevance to and therefore little impact on individual cognitive development during college. Greater understanding may come from estimating contextual and relative standing effects at levels of aggregation that are not only theoretically justifiable but also more proximal and directly related to student learning (for example, classrooms, peer groups, roommates). In short, the most informative multilevel analyses are likely to be those "based on theories in which the source and form of group effects are measured directly" (Burstein, 1980a, p. 207).

It may be, of course, that aggregate effects at almost any level are simply too psychologically remote (or too globally measured) to have important direct effects on student development. Instead, the major aggregate-level influences on student development in college may be indirect, transmitted through their shaping of the individual student's interaction with important agents of socialization on campus, such as peers and faculty.

Change Scores

A substantial amount of the more recent research on the influence of college has a developmental focus and attempts to estimate how exposure to different collegiate experiences or environments leads to differential change on some trait over time. For example, do students who reside on campus tend to change more in critical thinking than students who commute to campus? One way in which this type of question has traditionally been approached is to compare pre- to postdifferences (such as freshman-to-senior scores) on an appropriate measure between groups of interest. If, for example, students residing on campus tend to change more in critical thinking than do those commuting to campus, one might conclude that the residential experience increases the impact of college, at least on critical thinking.

This is an intuitively appealing approach. There are, however, two problems with the use of change scores: reliability and the fact that the magnitude of the change or gain is typically correlated with the initial score (Linn, 1986). Reliability is an issue because change scores incorporate the unreliability of both the pre- and posttest measures (Thorndike & Hagen, 1977). This can be a major problem when difference scores are used to make decisions about individuals, but it may not be a major issue when group comparisons are being made (Cronbach, 1970; Linn, 1986). The second problem with change scores, their correlation with the initial score, however, can confound attempts to attribute differential change to exposure to a particular group or educational experience. If one simply compared changes in critical thinking between residents and commuters, it would be extremely difficult, if not impossible, to determine whether the differential changes were due to differences in actual residence status or simply to differences in initial critical thinking status between the two groups.

Comparing simple change or gain scores cannot correct for the lack of random assignment to different groups or collegiate experiences. A better (though not totally adequate) approach would be to employ change or gain in critical thinking as a dependent measure in a regression model that includes both a measure of group membership (for example, 1 = residents, 0 = commuters) and initial level of critical thinking. This would indicate whether or not residence arrangement is significantly associated with critical thinking gains when the influence of initial critical thinking status is partialed out. It is of interest to note, however, that one need not use change scores to obtain essentially the same information. Exactly the same results in terms of the statistical significance of residence status would be obtained if senior-year critical thinking were regressed on a model containing both residence status and initial freshman-year critical thinking (Linn, 1986; Linn & Slinde, 1977). Similarly, in the fictitious example we have been using throughout this appendix, essentially the same *net associations* for *SIZE* and *IWF* would be obtained in either of the following regression equations:

$$IO_2 - IO_1 = IO_1 + SIZE + IWF + \text{error}$$
$$IO_2 = IO_1 + SIZE + IWF + \text{error}$$

In what has come to be regarded as a classic paper, Cronbach and Furby (1970, p. 80) have suggested that "investigators who ask questions regarding gain scores would ordinarily be better advised to frame their questions in other ways." In fact, as suggested above, questions about gain or change can typically be reformulated without sacrificing information. Regression analyses that treat the pretest (precollege) scores no differently from other independent variables in the model and use the posttest (senior-year) scores as the dependent variable provide essentially the same information while avoiding many of the problems associated with change scores (Linn & Slinde, 1977).

This is not to suggest that change should not be studied. Recent work by Bryk and Raudenbush (1987) and Willett (1988), as reviewed by Light, Singer, and Willett (1990), has suggested that the study of change becomes more valid and less ambiguous when it is measured over more than two time points. Light, Singer, and Willett (1990, p. 147) argue that assessing change over three, four, or even more time points permits one to trace the "shape of each student's growth trajectory" rather than just the difference between the beginning and end points. The use of multiple estimates of student status over time is a promising new methodological approach to the assessment of change or growth.

Conditional Versus General Effects

The analytical procedures we have discussed in the preceding sections have all assumed that the net effects of each independent variable on the dependent variable are general. That is, the effect is the same for all students irrespective of their status on other independent variables (Kerlinger & Pedhazur, 1973). Thus, in our fictitious example we are assuming that the net direct effect of IWF on IO_2 is the same regardless of the student's level on IO_1 or the size of the institution attended. This assumption certainly has the appeal of parsimony (that is, other things being equal, the simplest explanation is often the optimal one). On the other hand, it can be argued that assuming only general effects in one's analytical approach ignores individual differences among students attending the same institution or exposed to the same educational or instructional experience. These individual differences among students may interact with different institutional, instructional, curricular, or other educational experiences to produce "conditional" rather than general effects. In a conditional effect, the magnitude of the influence of certain educational experiences on the dependent measure may vary for students with different individual characteristics. Thus, for example, the magnitude of the direct effect of IWF on IO_2 may vary, depending upon the student's precollege level of intellectual orientation (IO_1) or on other individual traits such as gender or race.

It is also possible that there may be patterns of conditional relationships or interactions that involve different levels of aggregation (Bryk & Thum, 1989; Raudenbush & Bryk, 1988). In a contextual analysis, for example, individual aptitude may influence achievement differently depending upon the aggregate level of institutional, departmental, or residence unit aptitude. Similarly, there may be interactions among college experience variables that do not directly involve individual differences among student precollege characteristics. The influence of informal contact with faculty on intellectual orientation, for example, may vary in magnitude in institutions of different size enrollment. Conditional effects of the various types described above may be masked by analyses that consider general effects only. Under certain circumstances this may lead the researcher to conclude that the effects of specific educational experiences are trivial or nonsignificant when, in fact, they may have statistically significant and nontrivial influences for certain subgroups in the sample. Thus, a narrow focus on aggregate means or tendencies as an index of college impact may mask important changes in individuals or student subgroups. (See Clark, Heist, McConnell, Trow, & Yonge [1972] or Feldman & Newcomb [1969, pp. 53–58] for a more extensive discussion of this point.)

The concept of conditional effects determined by the interaction of individual differences among students with different methods of teaching or the presentation of course content has a respected tradition in instructional research. Here it is typically referred to as aptitude (or trait) × treatment interaction (Berliner & Cahen, 1973; Cronbach & Snow, 1977). Underlying its application in instructional research is the more general perspective, stemming from the psychology of individual differences, that not all individuals will benefit equally from the same educational experience. Applications of the investigation of conditional effects with postsecondary samples are provided by Romine, Davis, and Gehman (1970) for college environments and achievement; by Holland (1963) for career choice and academic achievement; by Pfeifer (1976) for race and grades; by Andrews (1981), Born, Gledhill, and Davis (1972), Buenz and Merrill (1968), Domino (1968), Daniels and Stevens (1976), Gay (1986), Horak and Horak (1982), Parent, Forward, Canter, and Mohling (1975), Pascarella (1978), Peterson (1979), Ross and Rakow (1981), and Stinard and Dolphin (1981) for different instructional approaches; by Cosgrove (1986) for the effects of programmatic interventions; by Bean (1985), Pascarella and Terenzini (1979a), and Terenzini, Pascarella, Theophilides, and Lorang (1985) in research on student persistence and withdrawal behavior in college; and by Chapman and Pascarella (1983) on students' levels of social and academic integration in college.

The computational procedure for estimating conditional effects involves the addition of a cross-product term to a general effects equation. Thus, if one is interested in the interaction of IO_1 and IWF, the required regression would be the following:

$$IO_2 = IO_1 + SIZE + IWF + (IO_1 \times IWF) + \text{error}$$

Because the cross-product of $IO_1 \times IWF$ is composed of variables already in the equation, its introduction produces a high level of multicolinearity or intercorrelation among the independent variables. Since this can lead to biased and unstable regression coefficients, the estimation of conditional effects is usually conducted via a hierarchical regression approach (Overall & Spiegel, 1969). In this approach, the general effects IO_1, $SIZE$, and IWF (sometimes called main effects) would be entered in the first step. This would be followed by the addition of the cross-product or interaction term in the second step. If the cross-product of $IO_1 \times IWF$ is not associated with a significant increase in R^2, one can then eliminate the cross-product term from the equation and interpret the equation in terms of its general effects results. If, however, the cross-product is associated with a significant increase in R^2, it suggests the presence of a significant conditional effect (that is, the magnitude of the influence of IWF on IO_2 varies with the student's precollege status on IO_1).

This being the case, the results yielded by the general effects equation would be misleading. Rather, one would interpret the nature of the $IO_1 \times IWF$ interaction to determine variations in the effects (unstandardized regression coefficient) of *IWF* on IO_2 at different levels of IO_1. Cohen and Cohen (1975) provide a simple computational formula for interpreting the nature of a conditional effect when the two interacting variables are continuous in nature. This formula can also be applied when one variable is categorical (for example, treatment versus control) and one is a continuous covariate (for example, aptitude). In the latter case an additional analysis can be conducted to determine the range of the continuous variable (aptitude) for which significant differences in the dependent variable exist between treatment and control groups (Johnson & Fay, 1950; Serlin & Levin, 1980).

A final point needs to be made about the estimation of conditional effects. The presence of replicable aptitude × treatment interaction effects has not been particularly common in experimental instructional research. Thus, in correlational data where one needs to rely on less effective statistical controls, the presence of conditional effects can often be artifacts idiosyncratic to the particular sample being analyzed. Considerable caution is therefore recommended in substantively interpreting conditional effects in correlational data. The most trustworthy are those suggested by theory and replicable across independent samples.

Final Note

At about the time this volume went into production, two potentially important books on the methodology of research and assessment in higher education were published. The first, by Light, Singer, and Willett (1990), uses case studies of actual investigations in postsecondary settings to introduce and explicate in greater detail many of the issues in research methodology touched upon in this appendix. The second, by Astin (1990), is a detailed treatment of many of the important conceptual, methodological, and analytical issues involved in assessing the impact of college and the impact of different experiences in college. Of particular relevance to the present discussion is Astin's own technical appendix on the statistical analysis of longitudinal data. Therein, he deals with many of the statistical and analytical issues we have just discussed, though from a somewhat different perspective. He also demonstrates how elements of regression analysis and causal modeling are combined to assess college effects within his input-environment-output model. Both books provide important conceptual, methodological, and analytical tools for scholars interested in the impact of college on students.