Researchers from various disciplines use techniques of investigation that are particularly effective in answering questions about specific populations. In this article, the author discusses one technique, path analysis, a model often used in the fields of sociology, economics, and political science.

STEVEN W. LECLAIR

Path Analysis: An Informal Introduction

In an attempt to maintain the "integrity" of one's own discipline, to communicate efficiently with colleagues, and to answer questions specific to each field, many scientific disciplines have established relatively exclusive research models. In most cases researchers from other areas have been unaware of these methods and techniques, unless some more-or-less formal collaboration or crossover was made between disciplines. In recent years, these coordinated efforts have become more popular and seem to be introducing many researchers to one another's techniques for use in their own investigations.

In the last two decades certain techniques for making causal inferences from nonexperimental data have been widely used in the fields of sociology, political science, and economics. I would like to present a simple discussion of one of these causal models called *path analysis*. The purpose of this paper is not to teach the reader how to use path analysis, nor is it to acquaint the reader with all of its complexities, but rather to offer an informal introduction so that the reader will become familiar with the techniques, strengths, weaknesses, and possible utility in counseling research.

Path analysis may be of particular interest and benefit to counseling research because of its interesting focus of providing possibilities for causal determinations among sets of variables. In other words, this technique tests a causal model developed from previous research or from a given theoretical stance and gives an indication of that model's appropriateness. This obviously has significant implications for research in the areas of career development, educational counseling, and (possibly) personal adjustment/problem-oriented counseling as well. In addition to its possible utility in the rather specific areas of individual and group counseling (where the ultimate focus is on one person's growth and development), path analysis may also hold benefit for program evaluation models in community, institutional, or school settings.

I wouldn't dare imply that this one technique is the answer to everyone's dreams in counseling research. Obviously, any technique is as strong as the individual using it, the research design used, and the theory on which the research question is based. Path analysis is simply another technique that can be added to the researcher's background of skills. It allows one to look at data in a manner that is somewhat different from traditional models and, if used correctly, can add to the knowledge we now have concerning how people and their environment interrelate. In the next sections I will discuss how path analysis relates to other methods of research, what it offers that others do not, certain limitations of the approach, and the level of research sophistication necessary to use it successfully. I will also introduce certain references that may be helpful for the reader to gain a more thorough understanding of the technique and its use.

BACKGROUND OF PATH ANALYSIS

The basic principles of path analysis were developed by a biologist, Sewall Wright, in the 1910s and 1920s (Heise, 1975). It was not developed to discover causes or causal relationships but to test the practical possibility of models developed by the researcher. Wright (1921) said: "In cases in which the causal relations are uncertain, the method can be used to find the logical consequences of any particular hypothesis in regard to them" (p. 557). It is important, then, for researchers to have a theory developed rather than relying on this causal inference technique to generate one for them. As this is the case, it is obvious that causal thinking plays a major role in the application of path analysis. To better understand what path analysis is, it is necessary to present a brief discussion of the concept of causation.

During my undergraduate and graduate training, I found that one of the favorite phrases of many professors teaching statistics and research design was: Correlation is no proof of causation. I agree with that statement and would go one step further to say that no index exists that is an ultimate proof of causation; however, many relationships, including correlations and covariations between variables, may be suggestive of causal linkages and patterns. Whether we are looking at the relationship between attitudes and perceptions or between traditional educational approaches and the degree of an individual's creativity, we are implying causation. Nagel (1965), in a discussion of the scientific community's view of causation, said, "Though the term may be absent the idea for which it stands continues to have wide currency" (p. 11). "The idea of cause," he continues, "is not as outmoded in modern science as is sometimes alleged" (p. 11).

At the crux of causal thinking, it can be said that one event, *A*, produces an expectation for the occurrence of a second event, *B*. A simple example of this relationship is exhibited in Figure 1.

The design in Figure 1 indicates that *A* implies *B*. The causal ordering, as can be seen in this illustration, is unidirectional ($A \rightarrow B$, not $B \rightarrow A$) and linear. Another possibility, however, is that *B* may be caused by events other than *A* and, in

STEVEN W. LECLAIR is an assistant professor, Faculty of Special Services, Ohio State University, Columbus.

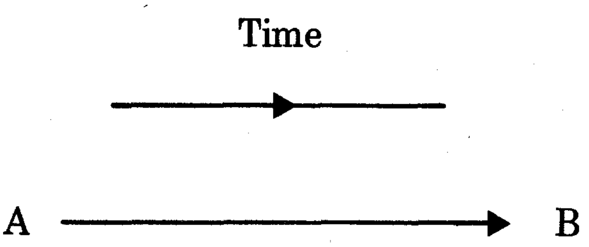

FIGURE 1

Illustration of a Two-variable Causal Relationship

those cases, *B* may occur in the absence of *A*. In other words, while the event *A* may imply the occurrence of event *B*, event *A* need not necessarily occur in order for event *B* to occur.

Figure 2 illustrates a few variations of three variable causal models. In Figure 2(a), *A* causes *B* and *B* in turn causes *C*. There is no direct causal connection from *A* to *C*. In Figure 2(b), *A* is a cause of *C* and *B* is a cause of *C* but there is no causal ordering inherent between *A* and *B*. In Figure 2(c), *A* causes *B* and *B* in turn causes *C*. In this case, however, *A* exhibits a direct causal influence on C as well.

There were some simple two- and three-variable illustrations of causal models. Obviously, it is possible for more than three variables to be present in the model involved in quite complex relationships. Causal models such as these can be developed but it must be remembered that the more variables one includes in a model, the more difficult it may become to interpret the results due to the interrelationships between variables. It must also be kept in mind that particular events in series of events are never universal causes of others. A causal model has the potential to occur only if the variables are suitably located in a system within which all conditions are "right" (Heise, 1975). For example, matches do not ordinarily cause forest fires. That causal relationship only exists when: (a) the matches are burning, and (b) they are introduced in some way to properly dry timber. This structural process beyond the specific variables in the causal relationship must exist for the theorized causal scheme to occur.

One other concept that is important to be aware of in causal thinking is that of a homogeneous flow. It is assumed that any event is a combination of lower level events that occur repeatedly as long as conditions remain constant.

Many cultural and social variables that seem to be categorical can be defined in process (flow) terms. For example, Heise (1975) states that sexual identity can be viewed as an ongoing presentation of specific types of cues in interaction. Sexual identity becomes a lifelong series of events that can be illustrated as a flow related to biology, early socialization, and present behavior. So the concept of flow is central to causal thinking in that it assists the researcher in conceptualizing all events as process.

Causal thinking implies the development of a unidirectional, linear explanatory scheme that, if it is to reliably occur, must be exposed to conditions that are recognized to aid its flow.

THE PROCEDURE OF PATH ANALYSIS

The statistical procedures involved with path analysis are basically methods of testing the appropriateness of a causal model with the use of standardized multiple regression equations. As with all nonexperimental methods, one of the major drawbacks is lack of control and subsequent inability to deal with all variables in a given system. The results can only be used as approximations to causality, so hopes for definitive causal explanations should not be entertained at this time (Miller, 1977). The power of the technique is impressing many individuals, however, and the frequency of its use is steadily growing.

Using traditional correlation and multiple regression techniques, the causal relationships among variables can only be inferred. These techniques often show evidence of important relationships, but when correctly using path analysis to test a theoretical model, it is possible to at least postulate causal linkages among a set of variables.

In order to "correctly" use path analysis, Kerlinger and Pedhazur (1973) incidate that the following four assumptions must be met: (a) relations among variables are linear, additive, and causal (curvilinear and multiplicative relationships and interactions are excluded); (b) residuals (to be introduced later) are not correlated; (c) a one-way causal flow exists; and (d) variables are measured on an interval scale.

Miller (1977) has developed a six-step sequence in the application of path analysis that will serve to illustrate its use.

1 Develop a causal scheme or model.
2 Establish a pattern of associations between the variables in the sequence.
3 Depict a path diagram.
4 Calculate path coefficients for the basic model.
5 Test for "goodness of fit" with the basic model.
6 Interpret the result. (p. 188)

STEP I: DEVELOP A CAUSAL SCHEME

The first step for the researcher is to develop a linear causal model based on a particular theoretical pattern or on previous research. It is particularly difficult to simplify the "real world," but an oversimplified model must be developed. The researcher must assume that the variables chosen are the most important, while all others not in the equation are referred to as "residual."

For purposes of illustration, why don't we assume that we are looking at the effects on job satisfaction of SES, IQ, and the level of educational attainment. Our hypothetical model may look like this:

(SES ⟷ IQ) ⟶ LEVEL OF EDUCATION ⟶ JOB SATISFACTION

Both SES and IQ imply the level of education attained, which in turn may infer the degree of job satisfaction. Another pos-

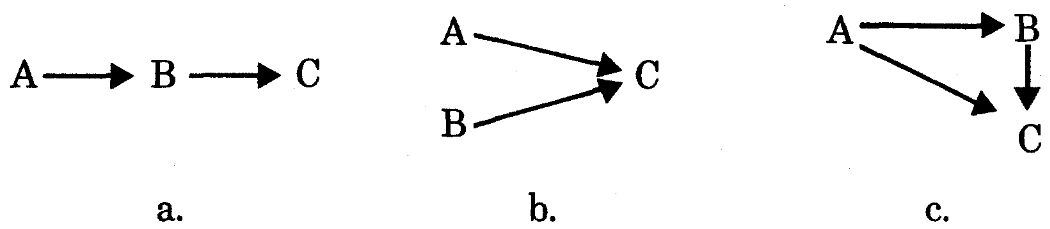

FIGURE 2

Illustration of a Three-variable Causal Model sibility in the model is that SES and IQ may imply job satisfaction not only through the intervening variable (Education) but also directly.

STEP II: ESTABLISH A PATTERN OF ASSOCIATION BETWEEN THE VARIABLES IN THE SEQUENCE

In this phase of the analysis, the scheme arrived at in the first step is put into quantitative terms. In order to do this a correlation matrix is developed by calculating the simple correlations between the four variables in the model. These simple correlations portray the general magnitude of effect that a prior variable has on a subsequent variable. The correlation matrix for our hypothetical example is illustrated in Table 1.

TABLE 1

Correlation Matrix for Four Variables

Variable		A	B	C	D
SES	A	–	.30	.41	.33
IQ	B	–	–	.16	.57
Education	C	–	–	–	.50
Job Satisfaction	D	–	–	–	–

STEP III: DEPICT A PATH DIAGRAM

A diagram of the causal model is useful but not necessary. It may be helpful in order to graphically display the interrelationships of the variables in a temporal sequence. An event whose variability is determined by other variables within the model is called an *endogenous* variable, while an event whose variability is explained by causes outside the model is called an *exogenous* variable. In our example, both SES and IQ are exogenous while level of education and job satisfaction are endogenous. Because it is impossible to account for 100% of the variability in the endogenous variables, the effect of all variables not in the model is called the residual. Figure 3 illustrates the path diagram of our hypothetical example.

Notice the inclusion of the residuals RES_1 and RES_2 to account for the proportion of variability not already explained in level of education and job satisfaction by the other variables in the equation.

STEP IV: CALCULATE PATH COEFFICIENTS

Kerlinger and Pedhazur (1973) say: "A path coefficient indicates the direct effect of a variable taken as a cause of a variable taken as an effect" (p. 310). To put it another way, a *path coefficient* reflects the total amount of variability in one variable explained by the causal impact of the other. Statistically, the path coefficients are identical to standardized multiple-regression coefficients (the betas). One method to obtain these statistics is to perform a multiple-regression analysis utilizing the causal scheme developed in Step I as a model for entering the variables into the equation. (I have provided a list of standard computer-regression programs at the end of this paper.) In Figure 4, the path coefficients have been added to the diagram.

TABLE 2

Path Coefficients and Coefficients of Determination

Dependent Variable	SES	IQ	Education	R^2
Job Satisfaction	.009	.501	–	.3527
Job Satisfaction	.398	.041	.416	.4966

STEP V: TEST FOR "GOODNESS OF FIT" WITH BASIC MODEL

Now that we have derived the path coefficients, we are interested in seeing how much variability in the dependent variable is explained by the variables as they exist in the causal model. For this, we look at the R^2, or the *coefficient of determination,* for each of the causal relationships we have designed. This allows us to assess the model's practical significance, or how much utility it may have in the "real world." Table 2 presents the path coefficients and the coefficients of determination for the causal model.

The results reported in Table 2 indicate that when SES and IQ alone are regressed on job satisfaction, they account for approximately 35% of the variability in the dependent variable. The additional amount of variance in job satisfaction determined when educational level is added is $R_1^2 - R_2^2$, or in this case, .4966 −.3527 = about 14.4%

STEP VI: INTERPRET THE RESULTS

The statistical significance of the analysis can be assessed by checking the *F* value derived in the multiple regression analysis with an *F* table under the appropriate degrees of freedom. Because the purpose of the analysis is to test the appropriateness of a particular causal scheme, however, it is important to identify the direct effect specific variables had on the dependent variable, and also the indirect effects each exerted through the causal scheme. It is also important to determine whether or not this causal scheme has practical significance so that it can be used with a fair degree of confidence. In our example it seems that close to 50% of the variability in Job Satisfaction can be explained by this model. Depending on how crucial the decisions are that will be made based on these data, it seems that 50% is a moderately strong indicator of practical utility. In view of the results, one could safely say that knowledge of

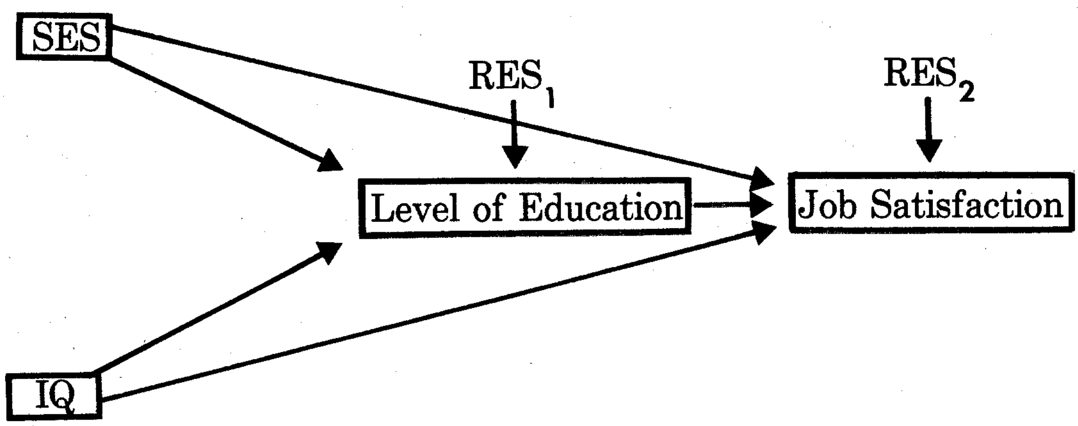

FIGURE 3

Illustration of a Path Diagram for the Four Variables

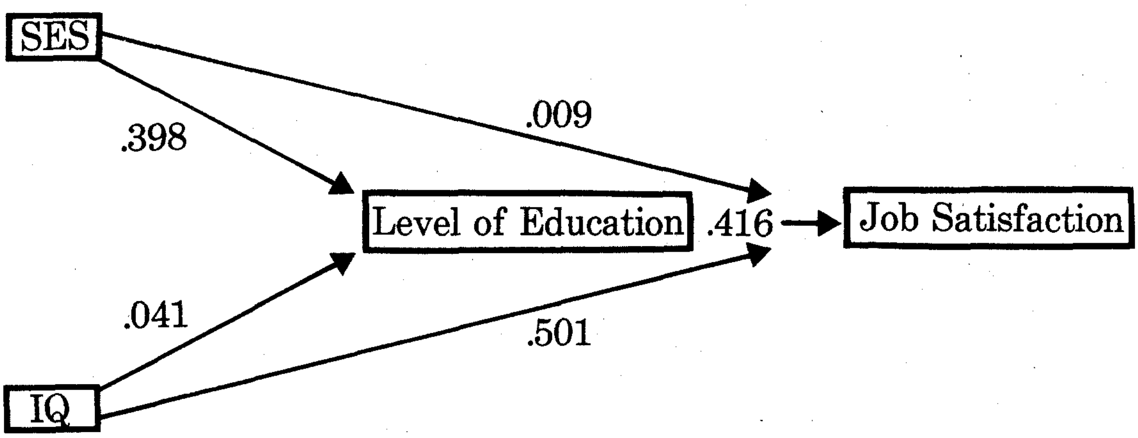

FIGURE 4

Presentation of Path Coefficients on Path Diagram

an individual's SES and level of educational achievement would enable you to quite successfully predict that person's degree of job satisfaction in the future. This result has significant implications for programs that are designed to enrich the lives of persons from lower-income and lower-social-status groups. The mission of many such programs seems consistent with the need of clients to attain more education and improve their income and social position. These results could be used to identify aspects of an individual's life that may be particularly sensitive with respect to job satisfaction. It must be remembered that this was purely a hypothetical example, and differences between this presentation and evidence from real data could be and probably would be significantly different. The example was simply used as a point of illustration.

SUMMARY

Path analysis, as presented in this paper, is a method of testing the appropriateness of a particular causal model that has been developed based on theory and previous research. It is primarily used in disciplines such as sociology, political science, and economics, but it has some exciting potential for use in the area of counseling research. To make use of this data-analysis technique, the researcher must be knowledgeable about the concept of causality and have practical expertise in multiple regression analysis with a capacity to conceptualize multivariate problems.

At the end of this paper is a short list of references that one could pursue in order to learn more of this technique and its applicability. As I said before, this has been an attempt at a crossover between disciplines in order to introduce counselors to a research method used in another field. Hopefully, it has served to interest a few readers and will help motivate them to use this technique and, possibly, to make an attempt at a similar crossover of their own. The answers to all of our research problems do not currently exist, but knowledge gained from investigating the systems of other fields may help to make our jobs a little easier.

REFERENCES

HEISE, D. R. *Causal analysis.* New York: Wiley, 1975.
KERLINGER, F. N., & PEDHAZUR, E. J. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.
MILLER, D. C. *Handbook of research design and social measurement.* New York: David McKay, 1977.
NAGEL, E. Types of causal explanation in science. In D. Lewis (Ed.), *Cause and effect.* New York: Free Press, 1965.
WRIGHT, S. Correlation and causation. *Journal of Agricultural Research,* 1921, *20,* 557–585.

ADDITIONAL READINGS

Books and Articles

BLALOCK, H. M. *Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press, 1964.
BLALOCK, H. M. *Causal models in the social sciences.* Chicago: Aldine, 1971.
DUNCAN, O. D. Path Analysis: Sociological examples. *American Journal of Sociology,* 1966, *72,* 1–16.
GOLDBERGER, A. S., & DUNCAN, O. D. *Structural equation models in the social sciences.* New York: Seminar Press, 1973.
LAND, K. C. Principles of path analysis. In E. F. Borgatta (Ed.), *Sociological methodology: 1969.* San Francisco: Jossey-Bass, 1969.
WRIGHT, S. The method of path coefficients. *Annals of Mathematical Statistics,* 1934, *5,* 161–215.

Computer Programs

DIXON, W. J. *BMPD: Biomedical computer program.* Berkeley: University of California Press, 1977.
NIE, N. H.; HULL, C. H.; JENKINS, J. G.; STEINBRUNNER, K.; & BENT, D. H. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.
SAS INSTITUTE. *SAS users guide: 1979 edition.* Raleigh, N.C.: SAS Institute, 1979.