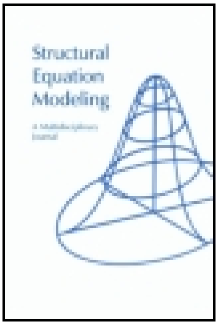

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Causation issues in structural equation modeling research

Heather E. Bullock [a], Lisa L. Harlow [b] & Stanley A. Mulaik [c]

[a] University of Rhode Island,

[b] Department of Psychology, University of Rhode Island, Kingston, RI, 02881-0808

[c] Georgia Institute of Technology,

Available online: 03 Nov 2009

STRUCTURAL EQUATION MODELING, 1(3), 253–267

Causation Issues in Structural Equation Modeling Research

Heather E. Bullock and Lisa L. Harlow
University of Rhode Island

Stanley A. Mulaik
Georgia Institute of Technology

As the use of structural equation modeling (SEM) has increased, confusion has grown concerning the correct use of and the conclusions that can be legitimately drawn from these methodologies. It appears that much of the controversy surrounding SEM is related to the degree of certainty with which causal statements can be drawn from these procedures. SEM is discussed in relation to the conditions necessary for providing causal evidence. Both the weaknesses and the strengths of SEM are examined. Although structural modeling cannot ensure that necessary causal conditions have been met, it is argued that SEM methods may offer the potential for tentative causal inferences to be drawn when used with carefully specified and controlled designs. Keeping in mind that no statistical methodology can in and of itself determine causality, specific guidelines are suggested to help researchers approach a potential for providing causal evidence with SEM procedures.

Increasing numbers of researchers in the social sciences have demonstrated a growing interest in path analysis and structural equation modeling (SEM; e.g., Jöreskog, 1970, 1973). Unfortunately, these methodologies have not always been appropriately applied. Much of the controversy surrounding SEM appears to be related to the degree of certainty with which causal statements can be drawn when using these methods. Despite cautions that correlation does not imply causation in the earliest statistics courses, re-

Requests for reprints should be sent to Lisa L. Harlow, Department of Psychology, University of Rhode Island, Kingston, RI 02881–0808.

searchers have been criticized for suspending this principle when employing SEM (e.g., Amdur, 1989; Games, 1988, 1990). Regardless of the increased sophistication and precision offered by SEM procedures, if the data are correlational in nature, no statistical method can change the design. Ultimately, it is the design, not the statistical method, that permits causal hypotheses to be adequately tested (e.g., Campbell & Stanley, 1963).

Although it is beyond the scope of this article to elaborate on all the conditions necessary to establish causality, three basic requirements have been cited: association between two variables (i.e., two variables must be correlated), isolation of the effect (i.e., ruling out extraneous variables), and temporal ordering, where a cause is shown to unambiguously precede an effect (for more discussion of these issues, see Bollen, 1989; Mulaik, 1986, 1987a, 1987b).

The purpose of this article is to examine the limitations as well as the strengths of SEM in relation to causality and to provide guidelines that may help researchers operationalize some of the pertinent issues. Our discussion is organized into the following sections: weaknesses and criticisms of SEM, strengths of SEM, and practical guidelines for concretizing some of the abstract prescriptions surrounding causality and SEM.

SEM: A CRITICAL LOOK

SEM methods have been applauded and attacked. Criticism ranges from relatively superficial to more substantive concerns about flawed logic (e.g., Baumrind, 1983; Cliff, 1983, 1987; de Leeuw, 1985; Freedman, 1987a, 1987b; Ling, 1982; Ragosa, 1987; Steiger, 1980). One must ask whether all of these criticisms are warranted.

Causal modeling (a name erroneously applied to SEM) has been criticized for containing the term *causal* in its name. Although seemingly superficial, the use of the term *causal* has been met with contempt. Guttman (1976) argued that causal analysis does not analyze causes or offer necessary or sufficient empirical conditions for testing causality. Bentler (1980) responded to this attack by explaining that *cause* denotes nothing more than the examination of a hypothesized, unobservable process and that terms such as *process* or *system* would aptly replace the term *causal.* Such a substitution might be beneficial. It appears that the term *causal* has led some researchers to incorrectly draw causal statements from these procedures. Commenting on this misuse, Guttman (1976) wryly stated, "Virtually every month current journals publish new causal analyses which undoubtedly puts sociology at the forefront of all sciences in terms of frequency of discovery of fundamental relationships" (p. 103). Although tongue-in-cheek, this statement effectively illustrates misconceptions about the degree of certainty with which causal statements can be drawn.

IGNORING BACKGROUND CONDITIONS

Background conditions such as mediating mechanisms, stability, and the form of the functional relation are often not adequately addressed in SEM applications (e.g., Mulaik, 1986, 1987b). Although it can never be determined whether background conditions have been fully met, researchers should make every attempt to ensure that such conditions have been considered.

A number of studies have examined various mediators to help explain a phenomenon. In the area of substance use, cigarette use and marijuana use have been cited as possible mediators between alcohol and harder substance use, offering support for a "gateway" hypothesis of substance use (Kandel, 1975; Welte & Barnes, 1985). In a different area, Guida, Ludlow, and Wilson (1985) found that "time on task" mediates anxiety and academic achievement, such that more time on task lessens the negative effects of anxiety on achievement. In still a different area, Harlow, Newcomb, and Bentler (1986) found that substance use and suicide ideation were mediated by a lack of purpose in life.

Stability refers to the need to measure the effect at the "correct" time interval. Making good estimates of the time interval can be difficult and often takes several studies to determine. This issue has been examined in a number of studies. For example, in a three-wave study of mediators of coronary heart disease, Fontana, Kerns, Rosenberg, and Colonese (1989) found that support was more stable 6 months later, whereas stress levels were more influential and stable after 12 months. In a different area, Francis, Fletcher, Maxwell, and Satz (1989) found that reading achievement was unstable across three waves of data from kindergarten to fifth grade, whereas cognitive skills remained fairly stable over time.

The *functional equation assumption* reminds researchers that the assumed form of a causal relation should be clearly stated (Mulaik, 1987a). Evidence should be accrued for each hypothesis over a number of different studies using different conditions to provide evidence that a causal mechanism is indicated. However, it is important to realize that countless functional relations that fit the same data may exist, and structural modelers should use caution and multiple sources of evidence to verify a specific functional relation. Kaplan (1989) pointed out that misspecifying a model may affect the power of z tests for other parameters in the model, highlighting the importance of careful specification of functional relations.

The extent to which background conditions are considered and met is the subject of sharp controversy. Freedman (1987a) insisted that all conditions that might weaken a causal inference must be shown to be inoperative before any causal statements are made. Freedman (1987b) claimed that the sheer complexity of SEM has led its users to be particularly lax in terms of attention to background conditions. Frequent use of SEM with nonexperimental data further complicates the ability to establish whether background conditions have been met. Although Freedman's advice is sound, it is virtually impossible for researchers to meet such requirements, even with experimental designs. Pragmatically speaking, researchers could approach this task by carefully considering the most likely defeaters of a causal hypothesis and then attempting to control these conditions. This is not too different from the imperatives advocated by Campbell and his colleagues (Campbell & Stanley, 1963; Cook & Campbell, 1979) to attend to threats to internal and external validity.

CROSS-SECTIONAL SEM STUDIES

To provide causal evidence, a minimal requirement is that a variable must be shown to temporally precede its effect. Temporal ordering may be difficult to establish in itself, particularly when working with structural equation models, because the latent variables cannot be observed. Unfortunately, cross-sectional applications predominate the literature, thus prohibiting evidence for directionality from being obtained. Longitudinal SEM procedures can offer some possibility of providing initial evidence for the direction of causation in a specific area of study. This preliminary evidence could be further verified with additional studies, preferably including experimental designs.

Anashensel and Huba (1983) presented a clear and understandable application of SEM with a longitudinal design examining depression, alcohol, and cigarette use across four waves. SEM methods for examining a mixture of cross-sectional and longitudinal data have been developed and may prove useful when data are not complete across all time spans of a study (McArdle, Hamagami, Elias, & Robbins, 1991). Farkas and Tetrick (1989) showed how findings can differ between cross-sectional studies and longitudinal designs, emphasizing the importance of examining several time points of data. Studies by Newcomb and his colleagues (e.g., Newcomb & Bentler, 1987; Newcomb, McCarthy, & Bentler, 1989; Stein, Newcomb, & Bentler, 1987) have demonstrated relations between substance use and other health and lifestyle variables measured up to 8 years later. Still another study, by Brook, Whiteman, P. Cohen, and Tanaka (1992), illustrates a three-wave longitudinal application on adolescent drug use that covers a span of 15 years. Although long-term data such as these are difficult to obtain, they provide much more evidence regarding the temporal sequence of variables and events than do cross-sectional designs.

CORRELATIONAL DATA

Much of the criticism of SEM stems from its most pervasive application with correlational data. In this case especially, the distinction between causal relations and causal cues must be constantly reinforced. In the strongest instance, correlational data can only provide cues to causal relations (e.g., White, 1990). The use of correlational data may be particularly treacherous when the researcher cannot actively manipulate variables to simplify relations, even temporarily (Cliff, 1983). Because SEM is frequently used to analyze nonexperimental data for which manipulation is impossible, this criticism is a valid one. However, SEM procedures can certainly be used with experimental data. Bagozzi and Yi (1989) suggested that SEM procedures be applied to experimental data that have traditionally been analyzed using multivariate analysis of variance (MANOVA) and multivariate analysis of covariance. It is important to reiterate that no statistical routine (e.g., MANOVA, SEM)—by itself—can establish causation; causal potential is determined by the degree of control and validity built into the research design. An example of this issue is found in Amdur's (1989) critique of causal models of delinquency. Amdur reanalyzed six studies that made causal claims with cross-sectional correlational data, pointing out conceptual and methodological problems with each. He concluded that much less is known about the causes of delinquency than has been claimed.

CONFIRMING A MODEL

Another problem concerns the use of the term *confirmed.* Some researchers seem to believe that confirmation of a structural model implies proof or exclusive validation of the model (Biddle & Marlin, 1987). Games (1988) drew a striking parallel between claiming support for a model and the faulty logic used when affirming the consequent. Unfortunately, evidence cannot always validate a model because it is possible that many other models may be equally acceptable. Garrison (1986) referred to this dilemma as the "undetermination of theory by experience" (p. 14). Because a potentially infinite number of models might be tailored to fit the data, the "best" model may remain underdetermined (Garrison, 1986). Although some researchers acknowledge that equivalent models may be equally consistent, many do not (see Breckler, 1990, for more on this concern). If two models account for the data equally well, in the absence of other considerations, the more parsimonious model or the one with the least number of parameters can be considered the superior model (see Bentler & Mooijaart, 1989, for a methodological rationale for this). In Breckler's (1990) review of 72 articles, only 1 acknowledged the existence of a specific equivalent model even though alternative models were plausible. The failure to identify alternative models may mislead some readers to assume that causal relations have been established. Therefore, confirming a model (i.e., retaining the null hypothesis) shows only that a model provides an "acceptable" description of the data (e.g., Biddle & Marlin, 1987). As the only legitimate statement that can be drawn, it is a far distance from a causal statement. Researchers are advised to compare a set of models when trying to establish important relations, instead of using a one-shot structural model assessment that most likely provides little evidence about causality, in and of itself. For an example of how models can be compared, see Huba, Wingard, and Bentler (1981). An excellent example of using two methods (SEM for correlation design, ANOVA for experimental design) to examine computerized training systems can be found in Coovert, Salas, and Ramakrishna (1992). Lawton, Kleban, Dean, Rajagopal, and Parmelee (1992) examined a confirmatory factor model over five different age-level samples to investigate a model of positive and negative affect over the life span.

LATENT VARIABLES

Bentler (1980) attributed part of the controversy surrounding SEM to its use of latent constructs that typically cannot be observed or directly measured. Latent-variable models introduce additional ambiguity to causal inference that directly observed variables do not (Mulaik, 1987b). Clearly, some researchers may be uneasy trying to establish causality through unobservable constructs.

Some confusion also stems from the possibility that what serves as one researcher's measurement model may serve as another researcher's structural model (P. Cohen, J. Cohen, Teresi, Marchi, & Velez, 1990). For example, P. Cohen et al. (1990) asserted that, although some constructs are clearly emergent and some are clearly underlying causes, the nature of others can be controversial. Socioeconomic status (SES) raises such controversy because it is unclear whether high income, prestigious occupation, and high educational achievement cause an individual's SES or vice versa (P. Cohen et al., 1990). For a discussion of the usefulness and the operationalization of latent variables, see Huba and Bentler (1982), Huba, Wingard, and Bentler (1981), and Martin (1982).

Latent and measured variables also raise issues surrounding what is referred to as the *nominalistic fallacy*—the fact that, even though we name something, we may not understand it (e.g., Cliff, 1983). In part, the nominalistic fallacy is an invalidity problem because the measured variables may be partially measuring something different from what we think they are measuring. The nominalistic fallacy may be particularly salient when considering models in which one or only a few measured variables are interpreted as defining a latent variable. Perhaps they do define a construct, but we can never be entirely certain what exactly is measured because latent constructs are "latent" by definition (Mulaik, 1987b).

As investigators, Cliff (1983) asserted that sometimes we reach into an incomplete grab bag of convenient variables, which are only suggestive of the nature of the true, underlying variables. This may be especially so when secondary analyses are conducted on data that were not intended for that specific use. High-quality data and well-thought-out models must be demanded (Cliff, 1987).

Taking a pragmatic stance, Martin (1987) asserted that most psychologists lack the resources to obtain the high-quality data on which to base latent constructs. However, uncertainty in defining latent variables may be reduced as the number of indicators and their individual validities increase (Cliff, 1983). Despite the desirability of using four or more measured variables per latent construct (Mulaik, 1987b), a review of 15 SEM applications revealed that 6 of the studies examined used only two indicators for at least one latent variable (P. Cohen et al., 1990). Cliff (1983) quickly pointed out that the status of a latent variable with three or four indicators, each with a correlation at .7, is still ambiguous. Even when four or more indicators are used, it is still highly possible that alternative sets of parameters would be equally consistent with the data and might lead to totally different conclusions concerning the nature of the latent variables. Passing the four-indicator test of a single common factor is a necessary, not a sufficient, condition that a single common factor has been identified.

Thus, two distinct issues arise when considering measured and latent variables: the difficulty of adequately assessing a latent construct, particularly when one measured variable is employed, and the difficulty of adequately naming a construct, regardless of the number of indicators that are used. It should be noted that both of these are validity issues that plague not only SEM but other statistical techniques as well. The effects of incorrectly specifying a latent variable can lead to the same sort of interpretational errors that are associated with an incorrect experimental manipulation (e.g., Tanaka, Panter, Winborne, & Huba, 1990).

POST HOC ADJUSTMENTS

Although post hoc adjustments are not limited to SEM, the proper role of ex *post facto should be briefly mentioned. It appears that many researchers* fiddle with parameters and variables after finding that a model is rejected. Although commonplace, this practice raises a host of both ethical and statistical criticisms. Considering causality, Mulaik (1987b) asserted,

> The potential objectivity of a model resides in its parameters fixed *a priori*. You can only test the objectivity of a model by testing fixed parameters against virgin data, that is, data that have never been used to determine the values of the fixed parameters of the model. (p. 30)

Otherwise, the researcher may be adjusting the hypotheses just so that they will conform to the data. In such cases, the model is not really tested because the data and the hypotheses would necessarily conform (Mulaik, 1990). Steiger (1990) emphasized the dangers of post hoc analyses without adequate statistical protection. Similarly, Cliff (1983) asserted that, after a model is adjusted in light of the data, the model loses its status as a hypothesis, and the model finally chosen represents a far less objective and stable representation. Thus, it would seem that potentially stronger causal statements might be drawn from a model that has not been adjusted than from one that has. If objectivity is a primary concern, this consideration should be kept in mind before adjustments are made.

FAILURE TO REPLICATE OR CROSS-VALIDATE

Many, if not most, SEM applications involve a single, one-shot model that may have had post hoc adjustments. This provides little or no opportunity to assemble causal evidence. If post hoc adjustments are made, however, researchers should replicate and/or cross-validate their findings on other data. Replication involves testing a model in a different sample under different conditions. Cross-validation is more stringent because it requires that the same parameter estimates from an initial sample be used in a second, independent sample. The extent to which the model fits the data with the restricted estimates provides some evidence of the generality of the model. Only portions of data not used in determining solutions for parameter estimates should be used to validate hypotheses because only these data will be independent in content from the hypotheses (Mulaik, 1990). Breckler (1990) suggested that researchers divide the original sample into two parts: a derivation sample and a cross-validation sample. The derivation sample could be used to fit the initial model and to develop modifications, whereas the cross-validation data could be used to assess the fit of the adjusted model (Breckler, 1990). Such cross-validation is essential when highly efficient computerized procedures—such as are used with SEM—may increase the likelihood that chance associations are processed as if they were real. Analyzing new data may clarify whether overfitting or fitting the model to random features has occurred (Bentler, 1978). Replication and cross-validation are always recommended unless the sample size is large enough to effectively represent the entire population with little distortion (Bentler, 1978). See Browne and Cudeck (1989) and Cudeck and Browne (1983) for a clear exposition of methods for cross-validating structural models. Numerous applications demonstrate possible replication and cross-validation strategies. Velicer, Huckel, and Hansen (1989) initially tested items in a first sample and then verified the structure in a second, independent sample. Mumford, Weeks, Harding, and Fleishman (1988) cross-validated performance in a derivation sample of 5,078 participants and a validation sample of 890 participants. Elliot's (1986, 1988) results varied and did not replicate across varying age and gender groups.

SEM: A "BRIGHTER" VIEW

Focusing on shortcomings gives an incomplete perspective of SEM. SEM does have strengths that may contribute to our ability to draw tentative causal inferences.

One benefit of SEM is that it allows latent constructs to be represented by multiple measures (e.g., Martin, 1987). This may be very advantageous for psychologists because it is unlikely that single measures can represent most major psychological constructs. Martin claimed that this is SEM's greatest asset. The use of multiple indicators may provide us with more valid and more reliable measurement of latent constructs. Latent variables can be considered a tradeoff. They raise issues surrounding the "nominalistic fallacy," but multiple indicators may provide accurately defined latent constructs and may reduce the severity of missing variables by providing more richly defined latent constructs.

Furthermore, using latent variables may allow researchers to use a limited number of exploratory constructs to explain phenomena. Philosophical justification for this use of latent variables is found in the works of both Thurstone and Kant. Thurstone advocated using a limited number of explanatory constructs common to a broad range of phenomena as a way of achieving objective constructs that are applicable to all phenomena, not just specific phenomena on an ad hoc basis. For a more complete analysis of the philosophical contributions of Thurstone and Kant to SEM, refer to Mulaik (1994, *in press*).

Second, Bentler (1980) asserted that SEM's great potential rests in its ability to handle very complex, multivariate models—particularly with quasi-experimental or nonexperimental research, in which methods for testing are not well developed. This may improve our ability to draw causal inferences because testing more sophisticated models and theories with good data and cross-validation may allow us to get a richer, more complete understanding of the phenomenon studied.

Third, even Cliff (1983) acknowledged that SEM is a powerful tool when it is used and interpreted properly. This is particularly salient when the aforementioned two benefits are combined. That is, SEM allows clustering and multiple-regression–type procedures to be performed simultaneously. Subsequently, inferences about the relation among latent constructs can be distinguished from any confounding effects both of error inherent in the measurement of the constructs and of variability in the items that are unrelated to constructs (e.g., Martin, 1987).

Fourth, the ability to allow the assessment of measurement error and prediction error is a great benefit that should not be overlooked. In this vein, SEM provides relatively unbiased parameter estimates among latent constructs that are each measured by two or more variables.

A fifth strength lies in SEM's ability to assess both direct and indirect effects (e.g., Biddle & Marlin, 1987). It is likely that more sophisticated models will include direct as well as indirect effects; SEM's ability to assess such complex models may provide analyses that are richer and more realistic and have the potential to reveal some of the complex causal relations that operate in our world. An interesting application on this was provided by Dembo et al. (1987), who showed that sexual abuse appears to directly affect substance abuse, whereas physical abuse tends to show both direct and indirect effects on substance use.

If causal evidence can be accrued through replication and cross-validation with rigorous designs and analyses, a procedure that allows multiple indicators and complex relations is necessary. Such a description aptly fits SEM.

SEM AND CAUSALITY: AN ASSESSMENT

In conclusion, does SEM offer anything "new" at the level of causal inference? Based on the previous discussion, the best answer appears to be both yes and no. In one sense, the answer is clearly negative because SEM cannot ensure that the necessary conditions of isolation, association, and direction of influence have been met. However, it should be noted that no other statistical procedure can ensure that these conditions have been met either. Although some could argue that SEM may not offer anything new at the level of causal inference, it does not offer anything less than other analyses.

On the other hand, the answer is affirmative because SEM does offer what can be referred to as compelling potential. The potential lies in SEM's ability to analyze direct and indirect effects, assess both measurement and prediction error, allow multiple measures to represent latent variables, and provide simultaneous estimation of measured and structural relations in a complex, integrated mathematical model. Although using latent variables may increase ambiguity, making causal inferences difficult, they also allow complex theories to be tested. Our world is a complex place, and, if causal evidence is ever to be effectively acquired, it will only be through designs and statistical procedures that can take such complexity into account.

Potential is the key word when discussing SEM. In our view, structural modeling does offer the potential to reach tentative causal statements. By continuously testing structural equation models, researchers may discover whether the causal hypotheses and functional equations on which their models are based are useful for explaining variables that are related to one another (James, Mulaik, & Brett, 1982, p. 96). However, it must always be remembered that SEM, as well as any other statistical procedure, offers no guarantees in terms of the necessary conditions. Researchers must constantly remind themselves that confirming a model in no way ensures its exclusiveness. Researchers forget that SEM, as well as all other statistical procedures, has the potential only to allude to causal relations. This forgetting can lead to incorrect conclusions about the causal nature of the data.

Unfortunately, the conditions necessary for establishing causation (isolation, association, and direction of influence) are ideals, and it is most unlikely that we will ever be able to determine whether these conditions have been met. This may leave researchers with a sense of hopelessness. As researchers, we are confronted with a dilemma. We can never realistically meet all necessary conditions; however, we cannot passively accept their inaccessibility. We advocate an active scientific approach that employs a number of rigorous procedures in an effort to begin to assess causality. In the absence of causal guarantees, Garrison (1986) encouraged researchers to practice the pragmatic virtues of epistemological conservatism and good sense. Researchers must actively rely and insist upon randomization, repetition, and replication, regardless of whether SEM or other analyses are used. Conclusions drawn from SEM must be drawn carefully and tenuously, just as with conclusions from other analyses. No analyses, alone, can directly prove causal relations. Researchers must remain critical of findings and not accept or draw causal conclusions simply because a sophisticated latent variable analysis was conducted (e.g., Martin, 1987). As SEM becomes more familiar, it is hoped that the confusion surrounding these procedures will diminish and that more emphasis will be placed on improving designs and the quality of data, thereby increasing the potential for building accurate causal evidence with SEM.

FINAL PRECAUTIONS AND GUIDELINES

By way of summary, several practical suggestions are offered to help researchers concretize some of the abstract prescriptions surrounding causality and SEM:

1. Regardless of the level of sophistication, no statistical procedure can ensure that the necessary conditions for establishing causality have been met. Be skeptical of any research that makes absolute causal claims. As with any other statistical procedure, SEM research must be evaluated in relation to the quality of the data, the status of the hypothesis relative to the theory, and the match between the substantive statement of the hypothesis and the design and statistical procedure used to test it (Martin, 1987).

2. SEM is frequently used to analyze nonexperimental data, and causal statements are difficult or impossible with nonexperimental data. However, even in the most well-designed experiments, one can never be certain whether background conditions have been controlled or whether all necessary conditions for establishing causality have been met. The ability to draw causal inferences cannot simply be determined on the basis of whether data are experimental or nonexperimental. It is also useful to consider what types of research questions are being asked before deciding what conclusions are legitimate. It is usually inappropriate to infer causal relations from purely exploratory or descriptive research.

3. It is impossible to know with absolute certainty whether a causal relation has been established. In the absence of such certainty, researchers must take every step possible to ensure valid conclusions. Researchers should:

a. Assess the relevance of and control for as many background conditions as possible.

b. Strive for longitudinal SEM designs to help assess the direction of causality.

c. Carefully operationalize latent variables.

d. Use four or more high-quality indicators per latent variable when it is appropriate.

e. Compare alternative models for a set of data.

f. Keep post hoc adjustments to a minimum.

g. Replicate and cross-validate all findings.

h. View each SEM study as just one part of a larger program of research to help understand a phenomenon. Other studies could include additional SEM and preferably one or more carefully planned experimental designs to further validate the findings. Certainly, results that hold over several studies using different designs provide stronger evidence than can be obtained from a single study. Several excellent studies that have incorporated a number of the aforementioned suggestions include Reynolds (1989) and Reynolds and Walberg (1991, 1992).

Following these general guidelines cannot guarantee that causal relations have been established with certainty. They do, however, increase the confidence with which tentative causal evidence can be accrued in a rigorous, ongoing scientific inquiry.

ACKNOWLEDGMENTS

This research was supported in part by National Institute of Mental Health Grant MH47233 to the University of Rhode Island (Principal Investigator Lisa L. Harlow) and by National Institute of Drug Abuse Grant DA01070 to the University of California at Los Angeles (UCLA; Principal Investigator Peter M. Bentler).

Thanks are extended to anonymous reviewers for helpful comments and criticisms and to Dr. Peter Bentler's laboratory for production assistance while Lisa L. Harlow was on sabbatical at UCLA.

REFERENCES

Amdur, R. L. (1989). Testing causal models of delinquency: A methodological critique. *Criminal Justice and Behavior, 16,* 35–62.

Anashensel, C. S., & Huba, G. J. (1983). Depression, alcohol use, and smoking over one year: A four-wave longitudinal causal model. *Journal of Abnormal Psychology, 92,* 134–150.

Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental design. *Journal of Marketing Research, 6,* 271–284.

Baumrind, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. *Journal of Personality and Social Psychology, 45,* 1289–1298.

Bentler, P. M. (1978). The interdependence of theory, methodology, and empirical data: Causal modeling as approach to construct validation. In D. B. Kandel (Ed.), *Longitudinal research on drug use* (pp. 267–302). New York: Wiley.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. In M. R. Rosenweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 31, pp. 419-456). Stanford, CA: Annual Review, Inc.

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106,* 315–317.

Biddle, B. J., & Marlin, M. M. (1987). Causality, confirmation, credulity, and structural equation modeling. *Child Development, 58,* 4–17.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin, 107,* 260–273.

Brook, J. S., Whiteman, M., Cohen, P., & Tanaka, J. S. (1992). Childhood precursors of adolescent drug use: A longitudinal analysis. *Genetic, Social, and General Psychology Monographs, 118,* 195–213.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445–455.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115–126.

Cliff, N. (1987). Comments on Professor Freedman's paper. *Journal of Educational Statistics, 12,* 158–160.

Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, N. C. (1990). Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement, 14,* 183–196.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings.* Boston: Houghton Mifflin.

Coovert, M. D., Salas, E., & Ramakrishna, K. (1992). The role of individual and system characteristics in computerized training systems. *Computers in Human Behavior, 8,* 335–352.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18,* 147–167.

de Leeuw, J. (1985). Review of books by Long, Everitt, Saris and Stronkhorst. *Psychometrika, 50,* 371–375.

Dembo, R., Dertke, M., laVoie, L., Borders, S., Washburn, M., & Schmeidler, J. (1987). Physical abuse, sexual victimization and illicit drug use: A structural analysis among high risk adolescents. *Journal of Adolescence, 10,* 13–34.

Elliot, G. C. (1986). Self-esteem and self-consistency: A theoretical and empirical link between two primary motivations. *Social Psychology Quarterly, 49,* 207–218.

Elliot, G. C. (1988). Gender differences in self-consistency: Evidence from an investigation of self-concept structure. *Journal of Youth and Adolescence, 17,* 41–57.

Farkas, A. J., & Tetrick, L. E. (1989). A three-wave longitudinal analysis of the causal ordering of satisfaction and commitment on turnover decisions. *Journal of Applied Psychology, 74,* 855–868.

Fontana, A. F., Kerns, R. D., Rosenberg, R. L., & Colonese, K. L. (1989). Support, stress, and recovery from coronary heart disease: A longitudinal causal model. *Health Psychology, 8,* 175–193.

Francis, D. J., Fletcher, J. M., Maxwell, S. E., & Satz, P. (1989). A structural model for developmental changes in the determinants of reading achievement. *Journal of Clinical Child Psychology, 18,* 44–51.

Freedman, D. A. (1987a). As others see us: A case study in path analysis. *Journal of Educational Statistics, 12,* 101–128.
Freedman, D. A. (1987b). A rejoinder on models, metaphors, and fables. *Journal of Educational Statistics, 12,* 206–223.
Games, P. A. (1988, April). Invited commentary: "Correlation and causation: An alternative view." *The Score, XI,* pp. 9–10.
Games, P. A. (1990). Correlation and causation: A logical snafu. *Journal of Experimental Education, 58,* 239–246.
Garrison, J. W. (1986). Some principles of postpositivistic philosophy of science. *Educational Researcher, 15,* 12–18.
Guida, F. V., Ludlow, L. H., & Wilson, M. (1985). The mediating effect of time-on-task on the academic anxiety/achievement interaction: A structural model. *Journal of Research and Development in Education, 19,* 21–26.
Guttman, L. (1976). What is not what in statistics. *The Statistician, 26,* 81–107.
Harlow, L. L., Newcomb, M. D., & Bentler, P. M. (1986). Depression, self-derogation, substance use, and suicide ideation: Lack of purpose in life as a mediational factor. *Journal of Clinical Psychology, 42,* 5–21.
Huba, G. J., Wingard, J. A., & Bentler, P. M. (1981). A comparison of two latent variable causal models for adolescent drug use. *Journal of Personality and Social Psychology, 40,* 180–193.
Huba, G. J., & Bentler, P. M. (1982). On the usefulness of latent variable causal modeling testing theories of naturally occurring events (including adolescent drug use): A rejoinder to Martin. *Journal of Personality and Social Psychology, 43,* 604–611.
James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.
Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika, 57,* 239–251.
Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Academic.
Kandel, D. B. (1975). Stages in adolescent involvement in drug use. *Science, 190,* 912–914.
Kaplan, D. (1989). A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research, 24,* 41–57.
Lawton, M. P., Kleban, M. H., Dean, J., Rajagopal, D., & Parmelee. (1992). The factorial generality of brief positive and negative affect measures. *Journal of Gerontology, 47,* 228–237.
Ling, R. F. (1982). [Review of *Correlation and causality*]. *Journal of the American Statistical Association, 77,* 489–491.
Martin, J. A. (1982). Application of structural modeling with latent variables to adolescent drug use: A reply to Huba, Wingard, and Bentler. *Journal of Personality and Social Psychology, 43,* 598–603.
Martin, J. A. (1987). Structural equation modeling: A guide for the perplexed. *Child Development, 58,* 33–37.
McArdle, J. J., Hamagami, F., Elias, M. F., & Robbins, M. A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research, 17,* 29–52.
Mulaik, S. A. (1986). Toward a synthesis of deterministic and probabilistic formulations of causal relations by the functional relations concepts. *Philosophy of Science, 53,* 313–332.
Mulaik, S. A. (1987a, March). *A conception of causality for structural modeling.* Paper presented at the Symposium on Methodological Issues in Causal Modeling, University of Alabama, Tuscaloosa.
Mulaik, S. A. (1987b). Toward a conception of causality applicable to experimentation and causal modeling. *Child Development, 58,* 18–32.
Mulaik, S. A. (1990, June). *An analysis of the conditions under which the estimation of parameters inflate goodness of fit indices as measures of model validity.* Paper presented at the meeting of the Psychometric Society, Princeton, NJ.

Mulaik, S. A. (1994). Kant, Wittgenstein, objectivity, and structural equations modelling. In C. R. Reynolds (Ed.), *Advances in cognitive assessment: An interdisciplinary approach* (pp. 209–236). New York: Plenum.

Mulaik, S. A. (in press). The critique of pure statistics: Artifact and objectivity in multivariate statistics. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 3). Greenwich, CT: JAI.

Mumford, M. D., Weeks, J. L., Harding, F. D., & Fleishman, E. A. (1988). Relations between student characteristics, course content, and training outcomes: An integrative modeling effort. *Journal of Applied Psychology, 73,* 443–456.

Newcomb, M. D., & Bentler, P. M. (1987). The impact of late adolescent substance use on young adult health status and utilization of health services: A structural-equation model over four years. *Social Science and Medicine, 24,* 71–82.

Newcomb, M. D., McCarthy, W. J., & Bentler, P. M. (1989). Cigarette smoking, academic lifestyle, and social impact efficacy: An eight-year study from early adolescence to young adulthood. *Journal of Applied Social Psychology, 19,* 251–281.

Ragosa, D. R. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12,* 206–223.

Reynolds, A. J. (1989). A structural model of first-grade outcomes for an urban, low socioeconomic status, minority population. *Journal of Educational Psychology, 81,* 594–603.

Reynolds, A. J., & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology, 83,* 97–107.

Reynolds, A. J., & Walberg, H. J. (1992). A structural model of science achievement and attitude: An extension to high school. *Journal of Educational Psychology, 84,* 371–382.

Steiger, J. H. (1980). [Review of *Correlation and causality*]. *Psychometrika, 45,* 403–405.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Stein, J. A., Newcomb, M. D., & Bentler, P. M. (1987). Personality and drug use: Reciprocal effects across four years. *Personality and Individual Differences, 8,* 419–430.

Tanaka, J. S., Panter, A. T., Winborne, W. C., & Huba, G. J. (1990). Theory testing in personality and social psychology with structural equation models: A primer in 20 questions. *Review of Personality and Social Psychology, 11,* 217–242.

Velicer, W. F., Huckel, L. H., & Hansen, C. E. (1989). A measurement model for measuring attitudes toward violence. *Personality and Social Psychology Bulletin, 15,* 349–364.

Welte, J. W., & Barnes, G. M. (1985). Alcohol: The gateway to other drug use among secondary-school students. *Journal of Youth and Adolescence, 14,* 487–498.

White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin, 107,* 3–18.