Problems in the Measurement of Latent Variables in Structural Equations Causal Models

Patricia Cohen
New York State Psychiatric Institute and Columbia University

Jacob Cohen
New York University

Jeanne Teresi and Margaret Marchi
New York State Psychiatric Institute

C. Noemi Velez
University of Puerto Rico

Some problems in the measurement of latent variables in structural equations causal models are presented, with examples from recent empirical studies. Latent variables that are theoretically the source of correlation among the empirical indicators are differentiated from unmeasured variables that are related to the empirical indicators for other reasons. It is pointed out that these should also be represented by different analytical models, and that much published research has treated this distinction as if it had no analytic consequences. The connection between this theoretical distinction and disattenuation effects in latent variable models is shown, and problems with these estimates are discussed. Finally, recommendations are made for decisions about whether and how to measure latent variables when manifest variables are potentially available. *Index terms: causal models, disattenuation, emergent variables, latent variable measurement, latent variables, structural equations modeling.*

The Goals of Latent Variable Causal Analyses

There is an increasing tendency to advocate the use of latent variable (LV) causal model analysis in the analysis of multivariate data (e.g., Bentler, 1980). This advocacy appears not only in publications in many areas of the behavioral sciences, but also in manuscript review and in peer review of grant proposals. There have been several researchers who have raised questions about the applicability of this methodology. Baumrind (1983) expressed doubts about the characteristics of the constructs used in the models. Cliff (1983) and Fornell (1983) warned about statistical and inferential practices often associated with their estimation.

In addition to purely statistical or purely theoretical matters, there are several other problems that lie at the intersection of statistical method and substantive theory with regard to LV models. Some of these problems have already been discussed with regard to path analyses of genetic models (Karlin, Cameron, & Chakraborty, 1983), and others have been identified more generally (de Leeuw, 1985; Fornell, 1983). This paper considers appropriate measurement of LVs, and uses real data and a review of recent published models for elaboration and illustration.

The purpose of LV analysis is to improve the accuracy and validity of inferences from empirical data. In order to accomplish this goal, many assumptions about the structure of the data and the meaning of the associations between variables must be made. Ideally, each of these assumptions will be based either on substantive theory or on the knowledge, derived from past empirical

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 14, No. 2, June 1990, pp. 183–196

0146-6216/90/020183-14$1.95

evidence, that certain assumptions may be violated without serious effect on substantive inferences. Unless these demands are clearly met, it may not be easy to decide whether the estimates provided by LV causal analysis can be considered to provide a superior basis for inference than alternative, less complex methods.

This discussion and survey of recently published articles will address three main issues, including:

1. The two general types of unmeasured operative variables,
2. The criteria for indicators of LVs, and
3. Disattenuation effects in LV models.

The final section proposes some guidelines for deciding whether and how to measure LVs in causal structural equations models. The discussion is deliberately confined to those research applications involving a test of causal models, because other applications—such as those testing for correspondence with a theoretically specified data structure, or those examining growth models—may be justified in datasets not meeting the criteria discussed here.

The Nature and Meaning of Operative Variables

The general class of operative variables (OVs) in structural equations (causal) models can be divided into two subsets—LVs and emergent variables (EVs). OVs are variables that are, in theory, causal or active because mechanisms for producing change in other variables are implicitly or explicitly posited by the theory (e.g., so that higher income may cause larger expenditures, higher parental education may cause higher offspring education, illness may cause pain, and alienation may cause distrust). In keeping with general definitions of the term *cause,* the discussion will identify variables acting in a causal or generative way (Baumrind, 1983; Holland, 1986) to produce changes in other variables. OVs may be single measured variables (MVs), or they may be LVs or EVs, measured by means of multiple MVs.

What is an LV, and how well do common constructs match an appropriate meaning? An LV can be considered to be a hypothetical construct (Cronbach & Meehl, 1955); yet, in theory, an LV cannot be measured directly, but rather in terms of those MVs upon which it has an effect (Bentler, 1980)—that is, by means of variables for which theory predicts that it has a causal influence. These empirical "stand-ins" for the unmeasured LV are therefore partially or entirely intercorrelated because of this unmeasured (LV) common cause. Thus these MVs are literally manifestations of or effects of the LV (in part), which is theoretically understood as the active force underlying or explaining their intercorrelations (Bentler, 1982). This causal model is represented and tested by a number of current LV model computer programs, including LISREL, EQS, and EZPATH.

However, the causal aspects of what is known as the measurement model are not always appropriately reflected in the above LV model representation. Some data are more consistent with the theory that MVs are causes of an OV that has not been (and perhaps cannot be) measured directly; or MVs may be related to the OV by virtue of unmeasured common causes. In these cases the variables might better be thought of as indicators, and the unmeasured OV may be called an EV to reflect its status with regard to its indicators. In such a case there is no reason why the EV would "explain" the correlations of the MVs that are used to measure it, nor is there any general reason why the MVs need to be intercorrelated at all. An example might be the components that are combined to produce the gross national product; a number of other examples actually used in the published literature will be given below.

MVs may, therefore, be used as stand-ins for unmeasured OVs. However, unmeasured OVs are of two types. The more familiar type is that implied in LISREL models, in which MVs are effects of OVs and are therefore referred to as LVs; this reflects their status as the active force underlying or explaining the correlations among the MVs with which they are connected (Bentler, 1982, p. 106).

The second type, EVs, are theoretically the causal link between the MVs and other variables in the model; however, they do not explain the correlations among their MVs, and theory does not even require that the MVs be positively correlated, or nonzero correlated. For example, the number of children in a family, illness of the mother, and hours of maternal employment are all valid and reliable indicators of a mother's availability to interact with and monitor any given child. Similarly, high serum cholesterol, high blood pressure, smoking, and family history of heart disease all cause vulnerability to heart attack and are indicators of cardiac risk. It is fairly easy to see that both mother's availability and cardiac risk are effects or emergent properties of their indicators and not their underlying causes. There would not be much disagreement in principle that cardiac risk and mother's availability are constructs that merely group the individual indicators together; they are simply superordinate labels for the MVs that may be thought of as generating the EVs.

When the MVs are in no reasonable sense caused by the OV, the observed correlations among them are not an appropriate basis for estimating their links with the OV. For example, if mother's illness and mother's employment are correlated, it is reasonable to assume that they have common causes or affect each other, but not that they are each caused by mother's available time for child-caring tasks (the proposed OV), which is clearly an effect and not a cause of these variables. The correlation between these variables is therefore irrelevant to the issue of their connection with the unobserved OV. These OVs may be measured as composites of MVs that are created prior to the causal analysis. As will be seen below, doing so will not provide the disattenuated estimates that are a major feature of LV causal analysis.

In factor analysis, depending on the purpose of the investigator, a factor can be viewed as falling at either end of a causal process—as an effect or emergent property at one end or as an underlying cause at the other (see Bentler, 1976, for a clear discussion of this distinction). In current computer programs that test structural LV models, however, LVs are generally required to be conceived as causes of MVs, regardless of the theoretical plausibility of this connection. Because the algorithm assumes that the correlations among MVs are caused by the LV (at least in part), the model estimates are not independent of this specification direction. Thus, in order for the model estimates to be valid, the specification must accurately represent the theory.

Some newly developed structural equations programs make it possible to estimate OVs that are emergent properties of indicators in at least some applications. However, it is often not possible to do so and retain an identified model, nor do the most popular computer programs allow for such specifications. Therefore, when causality in the measurement model is specified in the wrong direction, as when an LV is really an EV, the model effects are not ordinarily estimable. The model relies on the fact that an LV is the source or cause of the correlation between its indicators to obtain critical constraints on the estimates of the links between the latent and manifest variables, thereby producing uniquely identified estimates.

To demonstrate that this problem is a real one for current researchers, the authors surveyed the 15 articles using LVs in causal models that were published in 1985 or early 1986 in *The Journal of Personality and Social Psychology, American Sociological Review, The Journal of Gerontology*, and *The American Journal of Sociology* (Aneshensel & Yokopenic, 1985; Bachman & O'Malley, 1986; Cochran & Hammen, 1985; Huba & Bentler, 1984; Laumann, Knoke, & Kim, 1985; Liang, 1986; Lincoln & Kalleberg, 1985; Lorence & Mortimer, 1985; Naoi & Schooler, 1985; Neff, 1985; Newcomb, 1986; Newcomb, Huba, & Bentler, 1986; Pavelchak, Moreland, & Levine, 1986;

Piliavin, Thornton, Gartner, & Matsueda, 1986; Reisenzein, 1986). Table 1 provides some identified LVs that are more appropriately identified as EVs because logic would suggest that they are caused by, or are mere superordinate labels for, their MVs.

It is probably not that uncommon for a construct to be measured either by its causes or by its (functional) consequences. For example, physical ill health is an important variable to be included in many structural models involving human well-being. It refers to the presence and severity of conditions such as cerebrovascular and cardiovascular disease, muscular-skeletal disease, cancer, and immune system related problems. In the elderly it can be measured with a series of reliable and valid scales assessing each of these conditions. Obviously, though, ill health as an OV so measured is an emergent property or effect of the indicators. It must therefore be expected that these MVs will only be minimally intercorrelated, if they correlate at all.

An alternative way of measuring ill health would be to focus on the functional consequences of disease, such as pain severity and persistence, energy level, fatigue-proneness, and activity limitation. These may be more properly thought of as consequences or effects of a general health or illness LV. Thus it is quite appropriate that one investigator's measurement model may be another's structural model, and a construct inappropriately measured as an LV in one study may be appropriately measured in another.

Although some constructs are clearly emergent and some are clearly underlying causes, the nature of others is quite controversial. Socioeconomic status (SES) is one example. It is a matter of considerable debate whether high income, a prestigious occupation, and high educational attainment cause a person's high SES or vice-versa. Different elements of this group of variables are differentially important to different outcome variables (e.g., Kessler, 1982), and there are readily discriminable subdivisions of the concept

Table 1
Latent Variables and Indicators
With Inappropriate Causal Direction

Latent Variable	Indicators
Stressful change events	Been sexually attacked Family and parent stress Accident and illness events Family relocation events
Life change	Number of undesirable events Number of events producing at least moderate life change
Illness	Number of illnesses Respiratory problems or illnesses Cardiovascular or circulatory problems
Monitoring capacity of organization	Number of domain-specific staff Number of Washington staff Number of technical staff
Instrumental activity of daily living	Minor repairs of appliances Driving a car a short way Taking a train or airplane trip Planting and maintaining lawn

Note. Examples from Newcomb (1986); Neff (1985); Aneshensel and Yokopenic (1985); Laumann, Knoke, and Kim (1985); and Liang (1986), respectively.

(Stricker, 1978). Because children take on the SES of their parents, in some cases it seems suitable to think of a child's socioeconomic status as causing his or her educational, occupational, and income status. However, later in life, SES seems less like a cause than an emergent property of its own indicators.

If SES is thought of as a true LV, it is not unitary with regard to its effects on other dependent variables, to which sometimes education and at other times income will be most highly related. These changes in relationship will also alter the measurement model. Were SES to be truly measured by those variables for which it may reasonably be said to operate as a causal functional unity, the indicators would probably be variables such as social attitudes and values, residential choices, and the characteristics of friends and associates. In practice, though, these variables are more likely to be selected as separate LVs caused by SES than they are to be used as indicators of SES.

On the other hand, it is theoretically appropriate to consider some constructs to be LVs operating as functional unities that have the power to influence the observed indicators. Some examples are depression (when the indicators are self-reports of feeling states), general intelligence (when the indicators are subtests of an IQ test), and personality traits (when the indicators are self or other report of behavioral tendencies and preferences).

Selection of Indicators for LVs

The Quality of MVs

The LV literature provides very little discussion of the question of how to measure the MV indicators of LVs, even after theoretical concerns have been satisfied. Should the indicators always be scales, or are individual items adequate indicators? Should LVs be first-order factors—measured by items—or second-order factors—measured by scales—or does it really not matter? One recent discussion reviews the factor-analytic literature for guidance on selection of MVs (Anderson & Gerbing, 1988). These authors note the objections raised by Cattell (1973) to factors based on items, and suggest a reasonable solution: to use composites of several items each. On the other hand, a solution that is suitable for exploratory factor analysis may not be suitable for testing of a theoretically-specified model. When MVs are single items, or subsets of items such as those that might alternatively have been added and used as a single scale, what is gained by treating them as individual indicators in an LV model? This question is discussed below with attenuation effects.

A review of the content of the MVs used for LVs in the surveyed studies supported an impression that many, and perhaps most, researchers are not clear about what standards ought to be used in selecting indicators. Of 14 studies, 13 used items or single facts (e.g., income) as MVs for at least some LVs, but 7 studies also used scales as MVs for other LVs. In more than one case, alternative functions of the same information were used as indicators (e.g., the use of an item in continuous and dichotomized forms as the two MVs).

Clearly all of these theoretical issues regarding the choice of MVs are of general importance for the production of fruitful research and data analyses. However, because LV analysis provides so many different options for assigning meaning to covariances, it is even more sensitive to deficiencies in the empirical realization of theoretical models than other methods of data analysis.

Number of Indicators and Their Intercorrelations

It has been shown that it is highly desirable to have four or more MVs for each LV (Mulaik, 1982), and it is almost necessary to have at least three (Costner & Schoenberg, 1972), especially if the measurement model is to be identified independently of the structural model. Nevertheless, 6 of the 15 studies listed above had only 2 indicators for at least one LV, and only two studies had four or more MVs for each LV. The modal number of MVs per LV was three, and the maximum number was 10. When scales (as opposed to items) were used, the number of indicators was typically either two or three.

Mathematically an LV is adequately specified if its MVs are so strongly correlated that the LV explains a high proportion of the variance of each MV, for example, over 50% (Fornell & Larcker, 1981). When MVs are correlated trivially or not at all, the causal estimates may approach empirical underidentification—a situation in which it is not the lack of coefficients but their small size that makes a wide range of estimated structural coefficients fit about equally well.

However, it can happen that an entity causes indicators that are not highly correlated. The somatic symptoms of depression, such as poor appetite and weight loss, increased appetite and weight gain, insomnia, hypersomnia, psychomotor agitation and psychomotor retardation (American Psychiatric Association, 1980, p. 214), are not nearly as highly correlated as self-report measures of depressive feelings; nevertheless they are generally considered to be caused by depression and not vice-versa. Because the somatic symptoms of depression are not highly correlated in the population, depression measured in this way would not make a good LV in a structural equations model.

In reviewing the problems surrounding the interpretation of LVs, Cliff (1983), citing McDonald and Mulaik (1979) among others, provides the following rather stringent guidelines:

> ...the definition or interpretation of [an] LV (or both) only becomes less uncertain as (a) the number of indicators and (b) their individual validities (communalities) increase. Furthermore, it seems that the status of a latent variable with only three or four indicators, each of which correlate .7 or so with it, remains very ambiguous. (p. 122)

In order to compare published studies with these guidelines, estimates of average intercorrelation among manifest variables were computed. When either the correlation matrix or the standardized measurement model loadings were provided, the average correlation among indicators could be determined or estimated from the products of the loadings. For the 15 studies reviewed here, intercorrelations averaged between .4 and .5 when scales were used as indicators, and between .1 and .3 when items were used. In two studies the items were virtual restatements, and the intercorrelations were much larger. Common patterns with the three-indicator models were a single large loading and two quite modest loadings, or two moderately large loadings and a quite small loading. In the former case the correlations among indicators were quite small, and the indicator with the high loading was more strongly related to variables in the structural portion of the model. In the latter case, two variables were more strongly correlated with each other than either was with the third. In sum, this survey suggests that the current literature is a long way from compliance with the standards suggested by experts in the area concerning the number and intercorrelation of MVs.

In general, the studies reported took little note of differences in loadings of MVs. Because loadings were so often quite variable, it would be an exaggeration to view the measurements as confirmatory, because they confirmed, at best, only the consistency of the data with a model in which at least some variables had significant loadings. Even simple sampling variation will tend to orient LVs in varying directions, especially when correlations between MVs are modest. Thus to some extent, each study, and sometimes each longitudinal reassessment in a single study, measures different "true" LVs, even when the same MVs are used.

Disattenuation: Influences on the Estimates of the Reliability of the LVs as Measured

A Comparison With Classical Disattenuation Methods

What exactly is meant when it is said that the causal estimates of the LVs' effects on other variables in the model "may be estimated directly" (Bentler, 1980)? It is that the estimate is made for the unobserved, perfectly reliable LV, in contrast to the imperfect measure that would be represented by some linear combination of the observed variables, usually their sum or their sum following standardization. Because the difference between this combination and the LV is attributed to unmeasured and uncorrelated causes, it may be thought of as unreliability. When relationships between any given LV and any other observed variable or LV are compared to the same relationships based on a simple linear combination of the observed indicators, it will be seen that the former must always be larger, because they have been "disattenuated" to take this unreliability into account. Although this observation is well known to experts in the field, its implications have not been systematically applied in LV analysis, nor has this literature generally been linked to the literature on disattenuated coefficients. Consequently, it is not uncommon to find LVs measured in structural models with effective reliability estimates that would be considered unacceptable in more conventional analyses and reports.

That this disattenuation is directly analogous to the classical correction for attenuation first proposed around the turn of the century can be seen in the simple model shown in Figure 1, Case A. Here, as consistent with current practice (Loehlin, 1987), causal direction is indicated by arrows, LVs are enclosed in ovals, and MVs in rectangles. Lower case subscripted letters have been used to refer to MVs corresponding to an upper case LV with the same letter. As drawn, for the sake of clarity, the relationships could have been limited to a measurement model. However, the addition of another LV with structural effects from X and Y would leave the conceptual issues essentially unchanged, though it would complicate the mathematics. Arrows from unmeasured causes have also been omitted from all figures for simplicity.

To correct the correlation between any X and Y for unreliability in X, the observed correlation is divided by the square root of the reliability of X. In this case the reliability (r_{xx}) is estimated by the correlation between the two measures, x_1 and x_2, $r_{12} = .5$. The corrected value is thus .2/.707 = .283. In the "classic" mode the reliability of the composite sum of x_1 and x_2 would be estimated by applying the Spearman-Brown formula (derived from Gulliksen, 1950, p. 78, Equation 10):

$$r_{X_c X_c} = \frac{k\bar{r}_{ij}}{1 + (k - 1)\bar{r}_{ij}} \quad , \tag{1}$$

where k is the number of elements being added, $\bar{r}_{ij}$ is the mean correlation, and it is assumed that the standard deviations are equal.

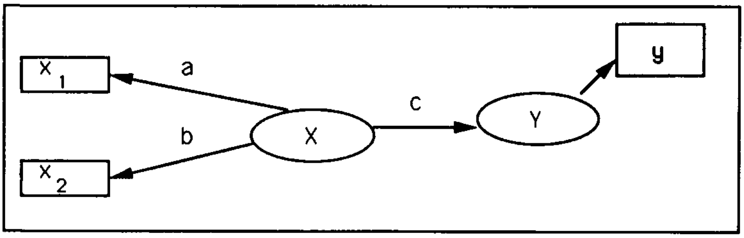
For $k = 2$, as in Case A in Figure 1, this simplifies to $2(r_{12})/(1 + r_{12})$, and for these data the reliability of the composite is $2(.5)/1.5 = .667$. The correlation between the composite and Y can be written as a function of the correlations (r_{iY} and r_{ij}) and k (derived from Gulliksen, 1950, p. 89, Equation 2):

$$r_{X_c Y} = \frac{\sum r_{iY}}{[k = k(k - 1)\ r_{ij}]^{1/2}} \quad . \tag{2}$$

For these data, the correlation of the composite with Y is $(.2 + .2)/[2 + 2(1)(.5)]^{.5} = .231$. When that value is corrected for attenuation by division by $.667^{.5}$, the result is .231/.816 = .283, the same .283 estimate obtained above (i.e., .2/.707 = .283). Now suppose there are four different x_i, all intercorrelating .5, and each with a .2 correlation with y (Figure 1, Case B). The LV causal estimate of .283 remains as before. Classically, from Equation 1, the reliability of the composite of the four x_i equals .8; and from Equation 2 the correlation of the composite with y is .253, which when corrected for attenuation in the composite is also .283.

Suppose now that the correlation in the two-indicator case is much smaller, say .2, although their correlations with y both remain at .2 (Figure 1, Case C). Under these new circumstances the estimated causal effect on y is $.2/.2^{.5} = .447$, a much larger value than obtained previously.

Figure 1
Causal Estimates and the Correction for Attenuation



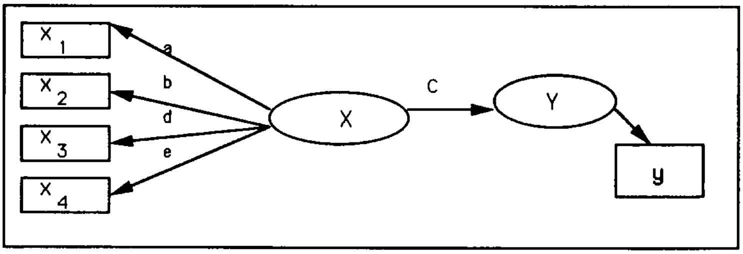
Case A. $r_{12} = .5$, $r_{yxi} = .2$, a=b= .707, c= .283

Case C. $r_{12} = .2$, $r_{yxi} = .2$, a=b=c = .447

Case D. $r_{12} = .5$, $r_{yxi} = .3$, $r_{y2} = .2$ a =.867, b = .577, c = .447

Case E. $r_{12} = .2$, $r_{y1} = .3$, $r_{y2} = .2$ a =.548, b = .365, c = .548

Case F. $r_{12} = .1$, $r_{y1} = .3$, $r_{y2} = .2$ a =.387, b = .258, c = .775



Case B. $r_{xij} = .5$, $r_{yxi} = .2$, a = b = d = e = .707, c = .283

As has long been known, disattenuation of relationships in multivariable models can result in either increases or decreases in the partialed relationships with other variables, including even changes in sign (Cohen & Cohen, 1983, pp. 408-409). Also, such estimated disattenuated standardized effects do not necessarily fall within the limits of ± 1.

Next assume that the observed correlations are as shown in Figure 1, Case D. The correlation of x_1 with y is .3, although that of x_2 is smaller (.2). The a and b estimates must still satisfy the constraint that $ab = .5$, the correlation between x_1 and x_2 in this just-identified model. It is still the case that $r_{Y1} = ac$ and $r_{Y2} = bc$, and the estimate of $c = .347$. Note that when the correlation between the indicators is smaller, the variable with the larger r with y dominates the estimate; see Figure 1, Cases E and F. Note particularly that a change in the structural portion of the model (r_{y1}) has affected the measurement model. This influence will be especially great when the correlations among the LV indicators are not strong, or when there are few indicators per LV, because the modeling program attempts to provide a maximally close fit to the original matrix of relationships, in this case a matrix in which there may be nearly as many structural coefficients as measurement coefficients.

For this reason, some writers have recommended that the measurement portion of the model be fixed prior to testing the structural portions of the model (Anderson & Gerbing, 1988), and one study reviewed actually followed this practice (Lincoln & Kalleberg, 1985). This practice is appropriate, but whether or not the measurement model is fixed, the effective reliability estimate should be reported. It is also possible that if such a practice were to be consistently followed, many findings would be shown to be substantially influenced by LVs for which the composite estimate is of low reliability—indeed so low that findings would have been questioned if more conventional methods had been used.

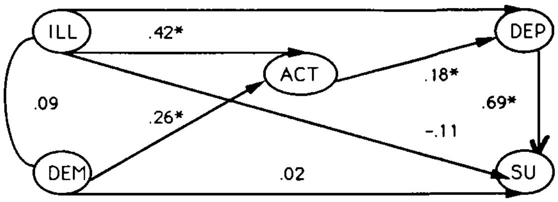
An Empirical Example

An LV causal model of the effects of physical problems, mental status, and activity limitation on depression and service utilization among the elderly was examined (Figure 2). The data came from a combined sample of 841 elderly community residents from the United States and the United Kingdom. Sampling strategies and characteristics of the sample are described elsewhere (Zubin & Gurland, 1977). Virtually all observed variables were internally consistent scales, with evidence of validity available from previous studies (see, e.g., Gurland, Kuriansky, Sharpe, Simon, Stiller, & Birkett, 1977; Teresi, Golden, Gurland, Wilder, & Bennett, 1984).

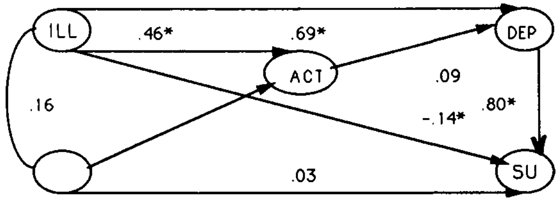
A number of researchers (e.g., Branch, Jette, Evashwick, Polansky, Rowe, & Diehr, 1981; Wolinsky & Coe, 1984) have found that health service utilization is predominantly accounted for by health and activity limitation. The geriatric literature reports strong relationships of both health and activity limitation with depression (e.g., Gurland et al., 1977). In the current model these constructs were measured with five and three MVs, respectively, and the respondent's mental status was included as a possible independent influence on depression and service utilization, each assessed with 3 MVs (Teresi, 1984).

Figure 2
Alternative Models of Illness, Depression, and Service Utilization Among the Elderly
(ACT = Activity Limitation, DEM = Dementia, DEP = Depression, ILL = Illness, SU = Service Utilization)

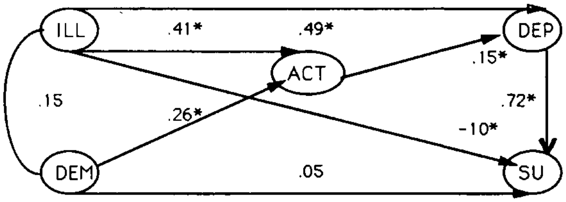
(a) Regression Analysis of Composites



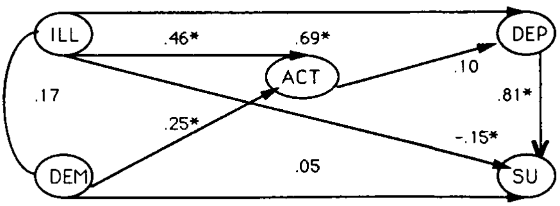
(b) Full LV Structural Model



(c) Regression Analysis of Estimated LVs



(d) Disattenuated Estimates for LVs
(ACT r_{xx} = .96, DEM r_{xx} = .96, DEP r_{xx} = .50, ILL r_{xx} = .67, SU r_{xx} = .67)



Estimated structural effects based on regression analyses of scale sums are provided in Figure 2a. In this analysis the effect of activity limitation on depression was statistically significant ($p < .01$). It can be seen from Figure 2b that the LISREL (Jöreskog & Sörbom, 1984) LV model estimate for the effect of activity limitation on depression was not statistically significant. The reason for the difference between these estimates may be seen in the composition of the illness LV. The five MV scales included problems with heart, stroke, cancer, and arthritis, respectively, as well as overall somatic symptoms regardless of disease. The illness OV can most reasonably be thought of as an effect of the disease indicators, and therefore as an EV with respect to these MVs, although it is a cause of the overall symptoms. As expected, each of the disease scales was correlated trivially or not at all with the others, and all were correlated moderately with the somatic symptoms. As a result, the illness OV was closely related to somatic symptoms (.89) and much less to each disease scale (average loading = .35). The three measures of activity level were more closely correlated, as were the dementia MVs.

Figure 2c gives the regression analysis of each construct as measured by weights generated from the LISREL solution. This weighted composite estimate of illness was more closely linked to depression, possibly because depression may also be manifested in somatic symptoms to some degree, a potential misspecification made more serious by the influence of the structural portion of the model on the measurement portion. However, even in this model the effect of activity limitation on depression was statistically significant.

When these LV model loadings were translated into reliability estimates, the estimate for illness was only .67. For activity limitation, in contrast, the more substantial correlations among the observed indicators led to a reliability estimate of .96. The reliabilities implicit in the LISREL analysis were applied to the weighted composites by the methods of classical disattenuation (Figure 2d). The result of differential reliability estimates was greater disattenuation of the illness effects on activity limitation and depression. The consequence was a much reduced and no longer significant effect of activity limitation on depression, once the estimated disattenuated causal effects of illness on both had been partialed. (These agree closely with the estimates in Figure 2b, confirming the above claim of equivalence between weighted disattenuated composites and LISREL solutions.) Do the results suggest that, contrary to theory and previous findings, it cannot be concluded that activity limitation leads to depression in the elderly, independent of coexisting illness? To accept the null hypothesis is, in effect, to accept the reliability estimate of the illness LV as appropriate, and, as has been seen, there is reason to believe that it is low—in part because illness is really an EV with respect to its (uncorrelated) disease MVs.

The Disattenuation Paradox

Because the knowledge of the consequences of unreliability for the magnitude of relationships between variables has long been known, why has the practice of disattenuation not been generally adopted (Block, 1963)? One reason is that the more poorly measured the variable is to start with, the larger the estimated disattenuated effects become relative to the observed effects, thus inflating sampling and other sources of error. The result is a lack of confidence in the superiority of estimates that take attenuation into account over those that do not. In real data the extent to which the correlations based on single relatively unreliable items are taken as the basis for strong causal inference could also be problematic.

An Alternative Approach to Disattenuated Estimates

There is an alternative that may solve some of the problems of OVs that prove inappropriate because of the nature or number of indicators. First, when internal consistency is not an appropriate measure of reliability, it is fully within the capability of current LV computer programs to accept any other estimate of the amount of measurement error provided by the analyst. In this case, indicators can be combined into a single measure in some theoretically correct manner. Then any appropriate estimate of the amount of error in the single measure may be specified in the input matrix, using test-retest or a simple educated guess, or model tests may be completed with both upper- and lower-bound estimates of the correlation of the measured EV with the hypothetical OV.

Prior specification of the reliability of composites would have two advantages. First, it would make possible the combination of more than one kind of measurement error—for example, error attributable to low internal consistency, if appropriate, and error attributable to imperfect temporal stability, if effects are hypothesized to occur over a longer period of time than is appropriate for the measurements. Second, it would make the magnitude of the disattenuation effects obvious to both the investigator and the reader of the research report. One article in the literature review appeared to have used this strategy, at least in part (Thornton, 1985).

The Realistic Limits of Current Theories

Loose Specification of Constructs

Measures of an LV are often chosen because they possess features like the term "depression" in the title, or because they appear to pinpoint different but important aspects of the construct under study.

Thus the choice of indicators is usually just as hypothetical as the causal hypothesis under investigation in the structural portion of the model. Also, because it is so speculative, the causal hypothesis is almost never clearly specified and therefore rarely comes under direct testing. Unless the LVs are solidly specified from a theoretical point of view, though, any causal relationships among them cannot be rationally assessed. Perhaps this loose specification that characterizes most theory accounts for some of the difficulties in publishing substantive LV models that many investigators have informally reported. When a method demands such complete specification, there is ample room for other investigators to disagree. One practical result of this dilemma is that little of the specification is provided. Some published reports seem to have taken this route.

Specification of Correlated Errors

The review of the literature indicated that omitting the presentation of correlated residuals among MVs is the rule, and virtually nowhere were correlated residuals discussed substantively, although the simple assertion that they were logically or theoretically reasonable did occasionally occur. Correlated error may be specified in observed variables that appear either as indicators of the same LV or of different LVs. However, the positing of correlated errors makes a theoretical statement—that is, the variance of the observed indicators is accounted for not only by the LV specified in the model and by random error (causes not shared with any other variable), but also by unknown shared causes. This implies, of course, the existence of yet another LV with the two (or more) observed variables as indicators. If researchers were required to identify these additional LVs, they would need to admit the limits of the theory being tested.

When single items are used as observed indicators there is usually very little hope of attaining such refinement in the specification. However, when indicators are scales selected on some theoretical basis, the correlation among their residuals may well be worth presenting. Doing so might, in particular, have the salutary effect of requiring careful consideration of what their common causes might be, especially in instances when observed variables with large positive correlations are estimated to have residual causes that operate in different directions (i.e., are negatively correlated).

Another potential problem comes from the usual inflexible requirement that all estimates be either in raw units or in standardized units. If the recommendation offered here is accepted—that items be combined into scales prior to use in an LV model—whatever standardization is required could take place prior to model estimation. The covariance matrix could then be the basis for the LV structural analysis. A useful mix of standardized and raw unit coefficients would be the result.

Implementation Problems

Almost every researcher consulted for this study reported consistent problems running models of realistic complexity. Examples considered for this paper included a study in which the theoretically appropriate and empirically available variables were far more numerous than could be included in the LV model. If LV models are to be measured with the recommended four or more variables each, there are few if any current programs that will produce estimates with more than seven or eight LVs. It is possible that these constraints are loosening; in any case, 25 or 30 observed variables would be the practical limit with the most popular programs. The resulting problems are particularly severe when the dataset is longitudinal and includes each measure at more than one point in time. When appropriately comprehensive models will not run, the overwhelming temptation is to eliminate the offending variables, even if it does not make theoretical sense; this is, or course, a truly Procrustean solution.

Recommendations for the Appropriate Measurement of Latent Variables in Causal Models, Including Reporting Practices

Given all the problems discussed above, the use of latent variable structural equations modeling should be approached with caution. Following are some guidelines for decisions about whether to use the technique, how to select indicators, and what information to include in reports of such work:

1. A strong theory, preferably with specific (and perhaps nested) alternatives, involving only relatively few latent variables should be available. Strong refers (1) to posited causes being truly generative, in the sense that the mechanism by which they could operate to produce differences in the affected variables could be explicitly stated; and (2) to clear expectations regarding causal effects that would include ranges of permissible correlations between residuals. This usually means that the method will be used primarily in those cases in which a "critical test" of a theory involving a modest number of constructs is being made.
2. LVs must have indicators of which they are plausible causes, and should have at least four well-correlated indicators. Much closer attention must be paid to the representativeness of indicators in order to assure that a theoretical construct (e.g., depression, SES) does not vary in its meaning in different studies. In general, the recommendation of Anderson & Gerbing (1988), among others, seems correct: The measurement model should be fixed prior to estimation of the structural portion of the model.
3. When only item level data are available as indicators, they should be combined prior to the analysis.
4. The reliability estimates of composites effectively employed by the LV models should be a standard part of the information provided in the report of the research.
5. It is also necessary to have studied univariate and bivariate distributions before venturing into more complex techniques. It may well be appropriate to implement the analyses with regression techniques, and to use LV analysis only after gaining a full understanding of findings not employing corrections for attenuation. In this way the effects of disattenuation and other features of LV modeling on the structural model may be explicitly and fully explored.

The proper use of LV models requires a combination of theoretical and statistical knowledge that is particularly difficult to obtain. Knowledge of the mathematical basis of these methods, though, does not ensure the appropriate test of a substantive theory. On the other hand, knowledge of the substantive theory and its most obvious and appropriate test does not ensure that the completed and reported test will either fit the theory or satisfy the mathematical demands of the statistical model. Most theoretically-sophisticated users will necessarily view the computer analysis as a sort of black box into which the data are fed, and out of which will come the best available conclusions about the substantive issues. If the attention conventionally paid to measurement in studies not using LVs were standard in studies using LVs, including the reliabilities, correspondence to measures used in other studies, and representation in the MVs of the theoretical domain, a higher level of agreement might be obtained among research results arising from differing data-analytic approaches.

References

American Psychiatric Association Committee on Nomenclature and Statistics (1980). *Diagnostic and statistical manual of mental disorders.* Washington DC: Author.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Aneshensel, C. S., & Yokopenic, P. A. (1985). Tests for the comparability of a causal model of depression under two conditions of interviewing. *Journal of Personality and Social Psychology, 49,* 1337–1348.

Bachman, J. G., & O'Malley, P. M. (1986). Self-concepts, self-esteem, and educational experiences: The frog pond revisited (again). *Journal of Personality and Social Psychology, 50,* 35–46.

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. (1976). Multistructural statistical models applied to factor analysis. *Multivariate Behavioral Research, 11,* 3–25.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31,* 419–456.

Bentler, P. M. (1982). Linear systems with multiple levels and types of latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Part 1, pp. 101–130). Amsterdam: North-Holland.

Block, J. (1963). The equivalence of measures and the correction for attenuation. *Psychological Bulletin, 60,* 152–156.

Branch, L., Jette, A., Evashwick, C., Polansky, M., Rowe, G., & Diehr, P. (1981). Toward understanding elders' health service utilization. *Journal of Community Health, 7,* 80–92.

Cattell, R. B. (1973). *Personality and mood by questionnaire.* San Francisco CA: Jossey-Bass.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115–128.

Cochran, S. D., & Hammen, C. L. (1985). Perceptions of stressful life events and depression: A test of attributional models. *Journal of Personality and Social Psychology, 48,* 1562–1571.

Cohen, J., & Cohen, P. (1983). *Multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale NJ: Erlbaum.

Costner, H. L., & Schoenberg, R. (1972). Diagnosing indicator ills in multiple indicator models. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences.* New York: Seminar Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

de Leeuw, J. (1985). Review of four books on covariance structure analysis. *Psychometrika, 50,* 371–375.

Fornell, C. (1983). Issues in the application of covariance structure analysis: A comment. *Journal of Consumer Research, 9,* 443–448.

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18,* 39–50.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Gurland, B. J., Kuriansky, J. B., Sharpe, L., Simon, R., Stiller, P., & Birkett, P. (1977). The Comprehensive Assessment and Referral Evaluation (CARE)—Rationale, development and reliability. *International Journal of Aging and Human Development, 8,* 9–42.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81,* 945–960.

Huba, G. J., & Bentler, P. M. (1984). Causal models of personality, peer culture characteristics, drug use, and criminal behaviors over a five-year span. In D. W. Goodwin, K. T. Van Dusen, & S. A. Mednick, (Eds.), *Longitudinal research in alcoholism.* Boston: Kluwer-Nijhoff.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationship by maximum likelihood, instrumental variables, and least squares methods.* Mooresville IN: Scientific Software.

Karlin, S., Cameron, E. C., & Chakraborty, R. (1983). Path analysis in genetic epidemiology: A critique. *American Journal of Human Genetics, 35,* 695–732.

Kessler, R. (1982). Disaggregation of the relationship between socioeconomic status and psychological distress. *American Sociological Review, 47,* 752–764.

Laumann, E. O., Knoke, D., & Kim, Y. (1985). An organizational approach to state policy formation: A comparative study of energy and health domains. *American Sociological Review, 50,* 1-19.

Liang, J. (1986). Self-reported physical health among aged adults. *Journal of Gerontology, 41,* 248-260.

Lincoln, J. R., & Kalleberg, A. L. (1985). Work organization and workforce commitment: A study of plants and employees in the U.S. and Japan. *American Sociological Review, 50,* 738-760.

Loehlin, J. (1987). *Factor, path, and structural analysis: An introduction to causal modeling with latent variables.* Hillsdale NJ: Erlbaum.

Lorence, J., & Mortimer, J. T. (1985). Job involvement through the life course: A panel study of three age groups. *American Sociological Review, 50,* 618-638.

McDonald, R. P., & Mulaik, S. (1979). Determinacy of common factors: A non-technical review. *Psychological Bulletin, 86,* 297-306.

Mulaik, S. (1982, November). *How should we overidentify structural equation models?* Paper presented at the meeting of the Society for Multivariate Experimental Psychology, Atlanta GA, U.S.A.

Naoi, A., & Schooler, C. (1985). Occupational conditions and psychological functioning in Japan. *American Journal of Sociology, 40,* 729-752.

Neff, J. A. (1985). Race and vulnerability to stress: An examination of differential vulnerability. *Journal of Personality and Social Psychology, 49,* 481-491.

Newcomb, M. D. (1986). Nuclear attitudes and reactions: Associations with depression, drug use, and quality of life. *Journal of Personality and Social Psychology, 50,* 906-920.

Newcomb, M. D., Huba, G. J., & Bentler, P. M. (1986). Determinants of sexual and dating behaviors among adolescents. *Journal of Personality and Social Psychology, 50,* 428-438.

Pavelchak, M. A., Moreland, R. L., & Levine, J. M. (1986). Effects of prior group memberships on subsequent reconnaissance activities. *Journal of Personality and Social Psychology, 50,* 56-66.

Piliavin, I., Thornton, C., Gartner, R., & Matsueda, R. L. (1986). Crime, deterrence, and rational choice. *American Sociological Review, 51,* 101-119.

Reisenzein, R. (1986). A structural equation analysis of Weiner's attribution-affect model of helping behavior. *Journal of Personality and Social Psychology, 50,* 1123-1133.

Stricker, L. (1978). "SES" indexes: What do they measure? *Basic and Applied Social Psychology, 1,* 93-101.

Teresi, J. A. (1984). *A comparison of the validity of scales developed using internal consistency, criterion-referenced and latent trait approaches to scale construction.* Unpublished doctoral dissertation, Columbia University.

Teresi, J. A., Golden, R. R., Gurland, B. J., Wilder, D., & Bennett, R. (1984). Construct validity of indicator-scales developed for the Comprehensive Assessment and Referral Evaluation interview schedule. *Journal of Gerontology, 39,* 147-157.

Thornton, A. (1985). Changing attitudes toward separation and divorce: Causes and consequences. *American Journal of Sociology, 790,* 856-872.

Wolinsky, F., & Coe, R. (1984). Physician and hospital utilization among noninstitutionalized elderly adults: An analysis of the Health Interview Survey. *Journal of Gerontology, 39,* 334-341.

Zubin, J., & Gurland, B. J. (1977). The United States-United Kingdom project on diagnosis of the mental disorders. *Annals of the New York Academy of Science, 285,* 676-686.

Acknowledgments

This work was supported in part by NIMH Grants No. MH36971 and MH3090, by the New York University faculty computer usage fund, and by the New York State Office of Mental Health. The authors thank Jeffrey Tanaka for his assistance in carrying out analyses, and Robert Golden for his suggestions on an earlier draft of the manuscript.

Author's Address

Send requests for reprints or further information to Patricia Cohen, Box 47, 722 West 168th Street, New York NY 10032, U.S.A.