Structural Equation Modeling in the Communication Sciences, 1995–2000

R. LANCE HOLBERT
University of Missouri-Columbia
MICHAEL T. STEPHENSON
Texas A&M University

Structural equation modeling (SEM) is a viable multivariate tool used by communication researchers for the past quarter century. Building off Cappella (1975) as well as McPhee and Babrow (1987), this study summarizes the use of this technique from 1995–2000 in 37 communication-based academic journals. We identify and critically assess 3 unique methods for testing structural relationships via SEM in terms of the specification, estimation, and evaluation of their respective structural equation models. We provide general guidelines for the use of SEM and make recommendations concerning latent variable models, sample size, reporting parameter estimates, model fit statistics, cross-sectional data, univariate normality, cross-validation, nonrecursive modeling, and the decomposition of effects (direct, indirect, and total).

Communication scholars have enlisted the statistical technique of structural equation modeling (SEM) for more than a quarter century, analyzing associations among a host of variables that exist at all levels of analysis. Cappella (1975) introduced the field of communication to the strengths, weaknesses, and assumptions of SEM and outlined how to construct and test a structural equation model. McPhee and Babrow (1987) then completed a critical assessment of the use of this technique in communication from 1976 through 1985, concluding "what our research community seems to have lacked is a clear format for the general execution and evaluation of path analyses" (p. 364). This study builds off and expands these works by analyzing the use of SEM in communication from 1995 through 2000.

R. Lance Holbert (Ph.D., University of Wisconsin-Madison, 2000) is an Assistant Professor in the Department of Communication, University of Missouri-Columbia, and *Michael T. Stephenson* (Ph.D., University of Kentucky, 1999) is an Assistant Professor in the Department of Speech Communication, Texas A&M University. The authors thank Robert M. Hauser, Department of Sociology, University of Wisconsin-Madison, and Rick H. Hoyle, Department of Psychology, University of Kentucky, for introducing them to SEM, and Michael Pfau and Philip Palmgreen for encouraging its use. Special thanks also to James Price Dillard and Edward L. Fink for providing some much needed points of clarification. Finally, the authors would also like to acknowledge the assistance of John McHale during the process of article collection. All correspondence should be addressed to the first author at 115 Switzler Hall, Columbia, MO 65211; email: HolbertR@missouri.edu.

Human Communication Research, Vol. 28 No. 4, October 2002 531–551

We focus on several important questions: Does a lack of unity concerning the formation and testing of structural equation models persist in the communication sciences? If so, what methods are used, and with what frequency are they employed? What methods are used to specify a model? How well do communication models conform to the basic assumptions of SEM? It is a necessary, but not sufficient, condition that these questions be addressed to help eliminate ambiguity and allow the field to advance in the use of this technique.

We use Baumgartner and Homburg's (1996) review of SEM in marketing and consumer research as a template from which to address the specification, testing, and evaluation of structural equation models in the communication sciences. This approach allows for both an assessment of the general use of this technique within a given field and a critical analysis of individual structural models. We also consulted several works that provide scholars with guidelines for evaluating (e.g., Hoyle, 1991) and reporting structural equation models (e.g., Boomsma, 2000; Hoyle & Panter, 1995; Raykov, Tomer, & Nesselroade, 1991). We begin with a general overview of SEM and then appraise the present state of SEM in communication. Recommendations are then put forth for where communication scholarship should turn its attention most immediately relative to SEM. Finally, we address some lingering concerns regarding the use of SEM and propose a framework to advance the use of this technique.

SEM

Overview

Largely developed through the work of Karl Jöreskog, present-day SEM is described by Kaplan (2000) as "a melding of factor analysis and path analysis into one comprehensive statistical methodology" (p. 3). Although the use of this technique has grown exponentially over the past two decades, its origins in path analysis are nearly a century old.

Jöreskog's SEM technique emerged from three separate lines of mathematical and statistical analysis: path analysis, factor analysis, and simultaneous equation modeling. Path analysis is commonly attributed to Sewell Wright (1934), who demonstrated how "correlations among variables could be related to the parameters of a model represented by a path diagram" (Kaplan, 2000, pp. 3–4). While Wright was refining his method of path analysis, Thurstone (1947) was advancing his approach to factor analysis (Muliak, 1972). Similarly, individuals in the field of econometrics began refining a technique of simultaneous equation modeling (see Blalock, 1971, pp. 153–157). Jöreskog's work bridged the gaps in these three areas. His earliest contribution, confirmatory factor analysis, was integrated with simultaneous equation modeling, resulting in the analytic framework we call SEM (e.g., Jöreskog, 1973).

SEM Fundamentals

Measurement and structural models. To understand SEM it is essential to grasp two fundamental concepts: the measurement and structural models. The measurement model establishes relationships between latent (unobserved) variables and multiple observable items. This is the confirmatory factor analysis portion of a model. Latent variables are the underlying constructs not directly tapped by any one set of measures, but they are hypothesized to influence certain observable items in the model. The latent variables are what a researcher ultimately wishes to capture, but which cannot be assessed directly through any one form of observation (Duncan, 1975).

The structural model tests a set of hypothesized associations among two or more variables. Although Jöreskog and others promote the testing of relationships among latent variables, as our analyses will reveal, many communication scientists have employed SEM to analyze associations among a set of observable variables (single-item or additive indices). The associations hypothesized among the variables (latent or observed) constitute the structural component of the model.

Boomsma (2000) encourages authors to provide diagrams of the structural and measurement models, error terms, and correlated parameters. In addition, Hoyle and Panter (1995) suggest that diagrams of the hypothesized and final models be presented, particularly if changes to the hypothesized model were made during estimation.

Covariance structure analysis. The statistical theory underlying SEM is grounded in covariance structure analysis, and the study of covariance matrices is preferred when using this technique (Cudeck, 1989). Authors should provide either the covariance or correlation matrices, and always include standard deviations for others interested in assessing their work (Boomsma, 2000; Hoyle & Panter, 1995; Rosenthal, 1984).

Model Specification

Approaches to modeling variables in SEM. As SEM has evolved, so too have the methods of analyzing relationships in structural models. We have identified three approaches to SEM within the communication sciences that contain a structural model component: observed variable (OV), latent composite (LC), and hybrid (HY).

The OV approach estimates relationships between observed (measured) variables only (e.g., Scheufele, 2000). OV models may employ single-item measures or composites and resemble the path analytic technique developed by Wright (1934). However, present-day OV modeling allows for full-information estimation not used in early path analysis.

The LC approach is similar to the OV technique in that it can employ composite or single-item measures. However, multiple-item indices are most common in LC models. Unlike the OV method, however, the LC approach specifies all variables as latent (e.g., Godbold & Pfau, 2000). The LC approach moves beyond the OV method by (a) specifying the composite as a single variable loading on a latent construct (all latent to observed paths are set at 1) and (b) fixing the error variance of this composite to (1 - reliability) times the variance of the indicator (see Bollen, 1989, p. 169). Unlike the OV approach, where measurement error remains within the variables, the LC approach accounts for unreliability by extracting measurement error from the latent constructs used in the structural model.

The third method is defined as a hybrid approach (Kline, 1998). In contrast to summing the measured variables to create a composite, the HY approach allows all measured variables to load individually on their respective latent variables (e.g., Fink & Chen, 1995). This approach is most consistent with the advances made by Jöreskog (1973).

Identification. All structural equation models must be overidentified (Hoyle, 1991; Raykov & Marcoulides, 2000). Overidentification occurs when there are more equations for the model than unknown parameters. Unknown parameters are those for which the SEM process will generate numerical values. If the model has one or more degrees of freedom (*df*), then the model is overidentified. If the model has zero *df*, then a model is just identified. The fit of a just-identified model is always perfect and cannot be properly evaluated (MacCallum, 1995). If a model is underidentified, most programs will not perform the analysis.

Recursive and nonrecursive models. Finally, structural equation models may be recursive or nonrecursive. Recursive models are most easily explained as models with unidirectional influences. By contrast, "nonrecursive models contain reciprocal causation, feedback loops, or they have correlated disturbances" (Bollen, 1989, p. 83). Caution must be taken when analyzing the variance accounted for in nonrecursive models (Teel, Bearden, & Sharma, 1986), and some have questioned the validity of testing reciprocal paths using cross-sectional data (Wong & Law, 1999). In short, nonrecursive models should be analyzed with care.

Model Estimation

Method of estimation. SEM users may select from several methods of estimation, including maximum likelihood (ML), unweighted least squares, generalized least squares, or asymptotic distribution free estimators (Jöreskog & Sörbom, 1996; Loehlin, 1998; Raykov & Marcoulides, 2000). However, the estimation procedures offered by various SEM software packages vary. Jöreskog (1973) proposed the use of ML to test structural equation models, and this estimator remains the most widely used (Bollen, 1989; Chou & Bentler, 1995).[1] ML is a large-sample estimator and assumes multivariate normality and thus normally distributed errors. Extensive testing of the ML estimator in Monte Carlo simulations suggests it is not biased with small samples provided that distributions are multivariate normal (Curran, West, & Finch, 1996), but other work in this area has identified problems with small samples that do not meet this requirement (Boomsma, 1983). When the sample size is large, ML is robust to deviations from normality (Hu, Bentler, & Kano, 1992). We do not recommend using ML with small samples that are multivariate nonnormally distributed as correct models are increasingly likely to be rejected.

Estimators fall into one of two categories, partial- and full-information. Because the approach advocated by Jöreskog (1973) builds on simultaneous equation modeling procedures, full-information estimators are recommended to acquire an optimal solution for a specified structural model. Full-information techniques are iterative procedures that estimate the entire system of equations specified in the structural model, while the partial- or limited-information techniques estimate one equation at a time. In analyzing systems of equations, the full-information procedures ascertain the influences on multiple dependent variables. In contrast, partial-information techniques, in analyzing one equation at a time, evaluate the influences on one dependent variable at a time. The full-information estimators isolate measurement and sampling error, thereby providing more accurate estimates of the relationships between variables. However, full-information is sensitive to model misspecification and "errors in any part of the system are more likely to ramify through the entire system of equations" (Blalock, 1971, p. 285). By contrast, error in partial-information estimation influences only the variables in the single equation being estimated.

Which estimator does one use? One should consider sample size, model complexity, and distribution of the variables before making a decision. Loehlin (1998) suggests running a model with multiple estimators. If the convergences are close, then the researcher should employ ML. However, if there are obvious violations to ML, then the asymptotic distribution free estimator is preferable. Authors should report the method of estimation and the distributional properties of their data. Boomsma (2000) suggests reporting descriptive statistics, including third and fourth moments (i.e., skewness and kurtosis). Finally, transformations to correct for nonnormality should be described.

Sample size. The statistical properties of the various estimators are dependent on large samples (MacCallum & Austin, 2000). Large sample studies ($N > 1000$), however, are not common in most existing communication research. Fortunately, this situation is not unique to a single discipline (e.g., Hoyle & Smith, 1994; MacCallum & Austin, 2000). As a result, researchers have focused on how small one's sample can be to still obtain replicable results.

Anderson and Gerbing (1988) recommend a minimum of 150 while Chou and Bentler (1995) suggest that a sample of 200 is "relatively small but practically reasonable" (p. 47). Tanaka (1987) suggests that more complex models require larger samples for stable estimates, although he detected fairly stable estimates in a sample with a 4:1 sample size-to-parameters ratio. In evaluating a three variable mediating model, Hoyle and Kenny (1999) found that parameter estimates were stable with a sample of 50, but only when latent variable measures were highly reliable ($\alpha > .90$). They recommend at least 100, but encourage 200 for simple mediating models with moderately reliable measures.

In the absence of a hard-and-fast rule, communication researchers should aim for a minimum of 150 participants per model. One should not rule out the use of SEM with fewer participants, but variables must be reliably measured, the structural models should be simple, and the limitations of the analyses must be documented.

Model Evaluation

Model fit. There are two types of fit indices: absolute and incremental (Bollen, 1989; Gerbing & Anderson, 1993; Hu & Bentler, 1995; Marsh, Balla, & McDonald, 1988). Absolute indices determine if the proposed model is consistent with the data without the use of a reference model. In contrast, incremental indices judge the "proportionate improvement in fit" by matching the hypothesized model with a nested baseline model (Hu & Bentler, 1995, p. 82).

The most common absolute fit index is the χ^2 goodness-of-fit test (Hoyle & Panter, 1995). Researchers desire a nonsignificant χ^2 to demonstrate that the specified model is not a null model. However, the χ^2-distributed test statistic is often problematic (Bentler, 1990). The χ^2-distributed test statistic is based on the statistical properties of T, and T is asymptotic. Therefore, T may not be χ^2 distributed with small samples (Hu & Bentler, 1995). Moreover, when a model lacks multivariate normality, the T may not be χ^2 distributed. With large samples, even trivial differences between a hypothesized model and sample data result in exponential increases in the χ^2-distributed test statistic. As there is increasing dissatisfaction with χ^2, many supplemental indices have been proposed to compensate for the ills associated with this test.

Jöreskog and Sörbom (1981) introduced their goodness-of-fit index (GFI) as an absolute test to counter the inherent weaknesses associated with the χ^2 test. However, Hu and Bentler (1998) detected numerous problems with the GFI and recommended that it not be reported. Hu and Bentler (1999) suggest reporting the standardized root mean squared residual (SRMR) in combination with one of the following incremental indexes: Tucker-Lewis index, Bollen 1989, relative noncentrality index, comparative fit index, or gamma hat.

Specifically, Hu and Bentler (1999) "recommend that practitioners use a cutoff value close to .95 for the Tucker-Lewis index (Bollen 1989, relative noncentrality index, comparative fit index, or gamma hat) in combination with a cutoff value close to .09 for SRMR to evaluate model fit" (p. 27). When sample sizes are below 250, the SRMR and one of the other recommended indices should be reported. When a sample is greater than 250, researchers may choose to combine SRMR with the root mean squared error of approximation (RMSEA; Browne & Cudeck, 1993). In this case, the SRMR should be close to .09 and the RMSEA close to .06 or less. While their recommendations are comprehensive, Hu and Bentler (1999) make room for practitioners to deviate slightly from these general guidelines (see pp. 27–28), although such deviations should be justified and clearly stated.

Because of their sensitivity to sample size, model misspecification, and distributional properties of the model's variables, Hu and Bentler (1999) determined the following fit indices perform poorly and should not be reported: goodness of fit index, adjusted goodness of fit index, normed fit index, Bollen 86, Akaike's information criterion (rescaled), cross-validation index, and critical *N*. Finally, authors should not report the χ^2 to degree of freedom ratio as an index of fit (cf. Kline, 1998). Marsh, Balla, and McDonald (1988) clarify that this ratio behaves mathematically similar to χ^2, and Bollen (1989) dismissed this ratio as unreasonable for assessing fit.

Parameter estimates and standard errors. Once researchers achieve a good model fit, they can examine the parameter estimates for the structural model. Hoyle and Panter (1995) recommend that all parameter estimates, their standard errors, and the variances of latent variables be reported. Authors should identify any post hoc model modifications made via LaGrange multiplier or Wald and report R^2 values for pertinent endogenous variables (Tanaka, 1993).

The issue of post hoc model modification is at the heart of a debate over whether SEM should be viewed as a strictly confirmatory method of analysis or as a technique that can also perform exploratory analyses. Jöreskog (1993) distinguishes between three types of modeling procedures: strictly confirmatory, alternative models, and model generating. The latter of these three types is the most exploratory, although researchers should have "at least some tentative ideas of what a suitable model should be" prior to beginning this procedure (p. 313). Others have argued that exploratory SEM procedures can too easily capitalize on chance during the process of final model construction, and past research finds that structural equation models constructed via various post hoc procedures are often not easily replicated (MacCallum, Roznowski, & Necowitz, 1992). Although we do not wish to take a firm stand on the issue of confirmatory versus exploratory SEM analyses, we do feel it is important that authors seek to better ensure the validity of their final models via cross validation. Cross-validation in SEM is akin to replication, and authors should report any attempt at cross-validation to a different sample (MacCallum, Roznowski, Mar, & Reith, 1994). An alternative is to report the single sample expected cross-validation index (Browne & Cudeck, 1993). The expected cross-validation index, computed from the existing sample, indicates how well that sample is likely to fit another independent sample.

Decomposition of effects. Three types of effects can be analyzed in SEM: direct, indirect, and total (Bollen, 1987). Direct effects are most commonly analyzed via SEM, and remain of primary interest to most research. Indirect effects assess the overall influence of one variable on another as that variable's influence works through one or more mediating variables (Hoyle & Kenny, 1999). In addition, indirect effects can be decomposed (i.e., specific indirect effects), allowing researchers to assess the level of mediation associated with particular variables within a model (Sobel, 1987). Lastly, the total effect of one variable on another is the sum of its direct and indirect effects.

Indirect effects are overlooked in most empirical research (Alwin & Hauser, 1975; Bollen, 1987). This is worrisome given that, "if an indirect effect does not receive proper attention, the relationship between two variables of interest may not be fully considered" (Raykov & Marcoulides, 2000, p. 7). Historically, various forms of communication are believed to have substantive indirect influence, and this is particularly true of mass communication (McQuail, 2000). Indeed, McGuire (1986) has argued that one of the ways in which empirical communication research can substantively advance knowledge is by analyzing indirect effects. Hence, we encourage communication scientists to not only evaluate the direct effects found in their models, but indirect and total effects as well.

METHOD

Data Collection

Each issue of 37 communication-related journals published from 1995–2000 was searched for the existence of SEM, resulting in a total of 59 articles.[2] Five percent of the issues were randomly selected after the initial search and reviewed by a third researcher. The third researcher was unable to uncover any works not obtained in the first search, providing further assurance that a valid census of articles was obtained.

Article Inclusion

This study focused on communication scientists' specification, estimation, and testing of structural relations among two or more variables (latent or observed) via SEM. Thus, the existence of the structure component of a structural equation model was a necessary condition for inclusion in this study. Studies that used SEM solely for confirmatory factor analysis were excluded from our analyses (e.g., Hackman, Ellis, Johnson, & Staley, 1999).

Several works employed a method of ordinary least squares regression path analysis (e.g., Hale, Lemieux, & Mongeau, 1995), but this technique is distinct from traditional SEM. These articles were not analyzed given their use of partial-information estimation (e.g., ordinary least squares), and the fact that this approach focuses on correlational rather than covariance analysis (e.g., Hunter & Hamilton, 1995). Additionally, one study conducted a two-stage least squares analysis (e.g., Davis, 1997), and this work was not included because this method is also a partial-information technique. Finally, McCutcheon and Nawojczyk (1995) used a latent logit path analysis procedure to analyze a two-group model. Although they used ML estimation, this method is defined as a "variant" of traditional latent variable models (p. 236). Thus, this article was also excluded.

Analyses

First, the article was treated as the unit of analysis, and assessments of publication year, frequency of use across journals, model type, type of data, and software package were made. Second, since 50% of the articles contained multiple models, the model was then treated as the unit of analysis. Each model was evaluated for initial specification, data screening, estimation procedures, testing, and methods of evaluation.

A total of 118 distinct structural equation models were found in the 59 articles. Some studies tested the same model across different samples. Other researchers tested different models on the same sample, and other works tested multiple models on multiple samples. We employed rules established by Baumgartner and Homburg (1996) to define distinct models. When the same model was tested across different samples, the data were averaged across replications. This created a single model that was then analyzed. For the other two cases (i.e., multiple models/same sample or multiple models/multiple samples), each model was treated as distinct.

RESULTS

General Trends in SEM

Frequency of use. An average of slightly less than 10 SEM articles per year appeared in communication journals from 1995–2000, with the greatest number of articles published in 1998 (N = 15). This volume is far less than that found by MacCallum and Austin (2000) in psychology for 1993–1997, but slightly greater than that identified by Baumgartner and Homburg (1996) in marketing and consumer research for 1977–1994.

Appearance of SEM within journals. Two communication journals distinguish themselves from the others in the publication of SEM studies, *Human Communication Research* and *Communication Research* (see note 2). These two journals accounted for 37% of the published SEM studies from 1995 to 2000, with 11 SEM pieces each. This is not surprising given that journals like *Human Communication Research* are recognized as being well advanced in terms of methodology and statistical applications (Emmers-Sommer & Allen, 1999).

Model type. The most prevalent type was the OV model, which treats all variables as observable. A little more than half (57.6%) of the papers collected for this study employed this model type. Roughly one third (35.6%) of the articles contained HY models. The LC models were the least used, accounting for 6.8% of the articles.

The prevalence of observable-only models is noteworthy given that SEM was created primarily to study relationships among latent variables (Bollen, 1989). The use of latent variables is generally regarded as the one true advancement of SEM above other multivariate techniques. Bentler (1980) states that the "greatest promise" for the advancement of social science is "multivariate analysis with latent variables and, more narrowly, linear structural equation (simultaneous equation, path analysis, structural relations, covariance structure) models with latent (unobservable, unmeasured) variables" (p. 420). The basic premise of Bentler's argument stems from Duncan (1975), who states:

> All observation is fallible, no matter how refined the measuring instrument and no matter how careful the procedure of applying it. In a strict sense, therefore, we never measure exactly the true variables discussed in our theories. In this same strict sense, all (true) variables are "unobserved." (p. 113)

In short, SEM affords researchers the opportunity to extract measurement error in order to analyze the true relationships among latent variables. In primarily using OV models, the majority of SEM work in the communication sciences does not take advantage of this opportunity. This results in an underutilization of one of the perceived strengths of SEM.

TABLE 1
Model Specification

	All models	*Observed variable*	*Latent composite*	*Hybrid*
	(N = 118)	(N = 60)	(N = 15)	(N = 43)
% Hypothesized model diagram	54.0	36.7[b]	46.7[c]	81.4
% Nonrecursive models	8.0	13.0	0.0	5.0
Sample size*	646.4 (37/4445)	332.91 (37/1253)[b]	194.07 (134/393)[c]	513.44 (84/1754)
% Sample size < 150	27.1	28.3[a]	73.3[c]	9.3
% Sample size > 1,000	10.2	15.0	0.0	7.0
% Discussion of model identification	7.6	10.0	0.0	7.0
% Evidence of data screening	16.1	11.7[a]	3.3[c]	9.3
% Distribution information	0.01	0.0	0.07	0.0
Number of degrees of freedom*	33.38 (1/454)	11.38 (1/43)[b]	4.73 (2/15)[c]	56.07 (15/191)
% Incorrect degrees of freedom	10.2	11.6	0.0	18.6

NOTE: Means or percentages provided where appropriate. High and low values in parentheses. * denotes where outliers have been discarded for OV, LC, and HY values. Scheffé post hoc tests: [a] = difference between OV and LC, $p < .05$; [b] = difference between OV and HY, $p < .05$; [c] = difference between LC and HY, $p < .05$.

Data. A vast majority (78%) of the SEM articles used cross-sectional data. Only 13 studies used longitudinal data. Special attention must be paid to the limitations of inferring causation from cross-sectional data (e.g., Cliff, 1983). The studies that use cross-sectional data may be better off envisioning their structural models as a set of relationships that can be viewed simultaneously, rather than arguing that they are establishing a set of purely causal relationships (Baumgartner & Homburg, 1996).

Software. LISREL was the most often used SEM software package among communication scholars (64%). EQS was used far less than LISREL (20%). Finally, AMOS and SAS Proc Calis were each used in two studies.

Model Specification

Hypothesized model. Only a slight majority of the hypothesized models were graphically displayed, with the HY models most often providing this presentational information (see Table 1). It is important that readers have a complete understanding of a proposed model, and one of the best ways to enhance reader comprehension is by providing a figure.

Nonrecursive models. Most structural equation models (92.0%) were recursive (see Table 1). Although these models developed directly from hypotheses and research questions outlined by the researchers, communication scholars should recognize that SEM does afford the possibility of testing reciprocal paths or feedback loops.

Sample size. The mean sample size for all models was 646.40 (Mdn = 311.50), with some models retaining a sample as small as 37 or as large as 4,445 (see Table 1). Those who employed the HY approach used larger samples than those who used either the OV or LC techniques. Moreover, 27% of the models had a sample size below 150, the minimum established by Anderson and Gerbing (1988). Special note must be made of the fact that an especially high percentage of the LC models fell below this threshold, but we also recognize that these models were fairly simple in terms of their number of estimated parameters. Conversely, only 10.2% of the models utilized a fairly large sample size (N >1000).

Model identification. There was little mention of proper identification (see Table 1). Most of the extant SEM literature which addresses the issue of identification focuses on HY models, and only 7.0% of the HY models were accompanied by a discussion of this kind.

Data screening. It is important for researchers to define how they dealt with missing values (Gold & Bentler, 2000). Overall, discussion of data screening occurred in only 16.1% of the models (see Table 1). Discussion of this kind was sparse with HY models, but more frequent with LC models.

Distribution information. Given that the estimation technique (i.e., ML) that dominates SEM studies assumes normality, there should be some discussion of variables' distributional properties beyond the traditional reporting of means and standard deviations (e.g., skewness and kurtosis). There was almost no discussion of this type of information for the models analyzed for this study (see Table 1).

Degrees of freedom. The average number of degrees of freedom (M = 33.38) indicates communication models were healthy in terms of overidentification (see Table 1). One of the more disturbing findings of this study is that roughly 1 in 10 of the reported *df* estimates did not match the model descriptions or what information the researchers provided graphically. If an accurate *df* count cannot be obtained, then something else is occurring in the model that is not being properly explained by the author(s).

Model Estimation and Evaluation

Input matrix. Proponents of SEM have long asked those who use this technique to provide the necessary information to allow other research-

TABLE 2
Model Estimation and Evaluation

	All models	*Observed variable*	*Latent composite*	*Hybrid*
	(N = 118)	(N = 60)	(N = 15)	(N = 43)
% Proper input matrix	21.2	21.7[a]	53.3[c]	9.3
% Identifying estimation procedure	39.0	38.3[a]	73.3[c]	27.9
% Parameter estimates	89.0	91.7	100.0	81.4
% Parameter standard errors	27.1	21.7	26.7	34.9
% Reliability of measurement	90.7	88.3	100.0	90.7
% Reporting variance accounted for	35.6	41.7[a]	0.0[c]	39.5
% Presentation of χ^2 value	83.1	73.3	100.0	90.7
% χ^2 significant	40.8	20.5[b]	0.0[c]	79.5
# Model fit statistics*	2.38 (0, 6)	2.51 (0, 5)[a]	0.93 (0, 1)[c]	2.70 (0, 6)
% No reported model fit statistics	10.2	11.7	6.7	9.3
% Goodness-of-fit index (GFI)	46.6	56.7[a]	20.0	41.8
% Comparative fit index (CFI)	38.1	23.3[a]	73.3[c]	46.5
% Evidence of respecification	42.4	40.0[a]	80.0[c]	32.6
% Cross-validation	4.2	6.7	0.0	2.3
% Indirect effects	14.4	21.7[a]	0.0	9.3
% Total effects	12.7	16.7	0.0	11.6

NOTE: Means or percentages provided where appropriate. * does not include χ^2 statistic. High and low values in parentheses. Scheffé post hoc tests: [a] = difference between OV and LC, $p < .05$; [b] = difference between OV and HY, $p < .05$; [c] = difference between LC and HY, $p < .05$.

ers to replicate their findings or test alternative hypotheses with the same data. Less than a quarter of the proper matrices (covariance or correlation matrix with standard deviations) were provided by the authors (see Table 2). Those who employed the LC approach were more likely to provide this information than researchers using the OV or HY model techniques.

Estimation procedure. Although it is common to assume that most models are tested via the ML estimation procedure (it is the default option for the various SEM software packages), researchers need to be explicit in identifying their method of estimation. This information was provided for a decent percentage of the models (39%), but the presentation of this information is not yet second nature (see Table 2). As with the presentation of an input matrix, the researchers using the LC approach tended to supply greater detail of how the data was analyzed.

Parameter and error estimates. An overwhelming percentage of the models contained parameter estimates (89%). In contrast, it was far less likely to find the error terms accompanying these path coefficients (27.1%, see Table 2).

Reliability. It is important that readers have some sense of the traditional reliability estimates of the measures being used in a particular study, no matter the model type. Communication scholars have adequately provided this information (see Table 2).

Variance accounted for. As opposed to reliability, the reporting of the variance accounted for in the criterion variable, or any other endogenous variables, was less common (see Table 2). Researchers provided some estimate of variance accounted for in 35.6% of the models, and the lack of this statistic was most apparent with LC models. The variance accounted for should not be treated as a test of the legitimacy of a model, given that a model can potentially fit without accounting for much variance (Bielby & Hauser, 1977). However, most scholars find it beneficial to know how much of the variance is accounted for as a result of a given set of associations.

Fit indices. The χ^2-distributed statistic was provided for a large majority (83.1%) of the models (see Table 2), and this was consistent across all three model types. However, none of the LC models and only 20.5% of the OV models produced a significant χ^2 value. By contrast, almost 80% of the HY models generated a significant χ^2. This may be due, in part, to the HY models utilizing larger sample sizes relative to the two other model types, with past research finding a "nearly perfect relationship" between sample size and model acceptability (Saris & Satorra, 1993, p. 182). However, there are most likely other factors causing the lack of fit as measured by the χ^2-distributed test statistic. Researchers must focus on the three assumptions associated with this likelihood-ratio test: multivariate normal distribution of observable variables, independence of observations, and a large enough sample size to benefit from asymptotic properties (Matsueda & Bielby, 1986). Otherwise, this fit statistic can fluctuate greatly as a result of even a minor deviation from one of these assumptions.

Communication researchers generally provided an adequate number of fit statistics (see Table 2). The GFI was reported for almost half of the models (46.6%), largely among those who employed the OV approach.

Given that the GFI is used extensively by communication scholars, it is important that we mention that a GFI of .90 or greater is required for indicating a solid model fit. However, it is also important to reemphasize that the GFI has been criticized as a poor measure of fit (Hu & Bentler, 1999), and we do not give our endorsement of this fit statistic.

The most popular incremental fit index was the comparative fit index, which was reported for more than a third of the models (see Table 2). Those who employed the LC technique typically reported this fit statistic. Once again, each fit index has been created to compensate for particular weaknesses in other statistics (i.e., compensation for nonnormality, small sample size, etc.) and researchers should choose only those fit statistics that match the characteristics of their data and model.

Respecification. Jöreskog (1993) outlined a process by which a researcher can move from a hypothesized model to a final model via the LaGrange multiplier test if the hypothesized model does not fit the data well. There was evidence that less than half (42.4%) of the models tested went through some respecification process (see Table 2). Most failed to identify the method used, but when a method was identified the LaGrange multiplier approach was most often employed.

Cross-validation. As a result of MacCallum et al.'s (1992) finding that the use of post-hoc respecification procedures can lead to a lack of model replicability, a call has been made for an assessment of cross-validation in all SEM analyses (Hoyle & Panter, 1995). There was almost no evidence of cross-validation of the models analyzed for this study (see Table 2).

Indirect and total effects. There was little discussion of indirect or total effects in the communication science SEM pieces (see Table 2). In only 14.4% of the models did researchers discuss indirect effects, and the discussion of total effects was almost nonexistent. Although the analyses of indirect and total effects are only as important as the hypotheses and research questions being posited, empirical communication research should recognize that with SEM comes the ability to assess more than just a set of direct relationships.

DISCUSSION

As with McPhee and Babrow (1987), we too find the lack of "a clear format for the general execution and evaluation of path analyses" in the communication sciences (p. 364). We identified three distinct methods for testing structural relationships in SEM. The major distinction between these three model types is how they deal with measurement error. In analyzing relations among only observable variables, (the approach most often used in the discipline) measurement error directly and adversely influences the estimated parameters. In contrast, the LC and HY models isolate measurement error by analyzing the structural relationships among latent variables.

We do not wish to recommend which SEM approach should be utilized in communication research, largely because the issue has not been reconciled among SEM practitioners (see discussion of measurement error in Bollen & Lennox, 1991). However, SEM users and their audiences should recognize that each technique has the potential to yield a distinct set of results. That is, each technique may produce different parameter estimates for the same structural model using the same data. Our own research demonstrates that OV parameter estimates are suppressed relative to the other two model types (Stephenson & Holbert, 2001). We would like to reemphasize that the work of Jöreskog and colleagues has been in the advancement of latent variable models. In employing the OV technique, the majority of SEM research in communication runs in opposition to the procedures commonly outlined in today's SEM literature (e.g., Bollen, 1989). Although observable-only models have their place in SEM, we would like to see greater movement toward latent variable models.

We wish to highlight five of the biggest concerns with recent communication-based SEM analyses. First, sample size is problematic. Although no floor has been established, we agree with Anderson and Gerbing (1988) that a sample of 150 is a conservative, but fair, standard to set for SEM analyses. More than one fourth of the SEM models analyzed fell below this level.

Second, we are concerned with the inaccuracy of the reported degrees of freedom. The reported *df* estimates could not be accurately obtained in more than 10% of the models. We encourage journal editors and reviewers to compare the reported degrees of freedom with what can be derived from the information provided in a manuscript (see Hoyle, 1991). In addition, it might be useful if authors asked colleagues to review a manuscript prior to submission for the purpose of checking degrees of freedom.

Third, the lack of any consistency in estimating model fit is troublesome. Everyone should report the χ^2 estimate given that it is still the best means by which to make comparisons across models (Hoyle & Panter, 1995). Perhaps more important, providing the χ^2 estimate best insures the reporting of degrees of freedom. Additionally, we recommend researchers report the SRMR as a second absolute fit statistic (e.g., Hu & Bentler, 1999). Authors should take into account sample size prior to reporting additional absolute or incremental fit statistics.

Fourth, the nature of the data is also of concern. Cross-sectional data were used in an overwhelming majority of the studies analyzed for this review. Communication researchers must recognize the inherent limitations associated with this type of data. There should also be a greater effort to report, at a minimum, additional univariate distribution estimates beyond means and standard deviations (e.g., skewness and kurtosis). We were unable to determine whether the normality restrictions inherent to ML SEM were problematic because these characteristics were rarely reported.

Fifth, communication scholars should make a concerted effort to try to cross-validate their models. Almost half of the models analyzed went through some process of model respecification. Many of the discussions concerning model modification failed to provide enough detail to allow for an understanding of the exact processes enacted in moving from a hypothesized model to a final model. SEM practitioners should detail any respecification of their models. Further, researchers should cross-validate models across data sets to best guarantee they are not capitalizing on chance when enacting various modification processes.

Finally, there are two components of SEM that communication researchers have not taken advantage of to date. The first component concerns the testing of nonrecursive relationships. Theoretical arguments for reciprocal relationships exist within the discipline (see DeFleur & Ball-Rokeach, 1989, for discussion of media dependency; Palmgreen, 1984, on media gratifications sought and obtained), but very few communication-based SEM pieces have tested nonrecursive models. Given that nonrecursive relationships abound within the discipline, researchers should begin to explore models that allow for an evaluation of these substantive theoretical questions.

The second component deals with the need to move beyond analyzing direct effects. Indirect effects are important in the study of communication (e.g., McGuire, 1986), and some of the dominant paradigms of media effects research (i.e., limited effects, conditional effects) acknowledge that the true role of media is most likely not a pure direct effect (McLeod, Kosicki, & Pan, 1996). Communication researchers need to hypothesize, test, and evaluate all three types of effects—direct, indirect, and total. Only by analyzing all three types will we gain a better understanding of the effects of all forms of communication at various levels of analysis.

NOTES

1. Jöreskog and others have also argued for the use of other estimation procedures (e.g., generalized least squares) so as to not be beholden to the strict assumptions of multivariate normality inherent to ML estimation (Browne, 1982; Jöreskog & Goldberger, 1972).

2. A bibliography of the articles analyzed for this study is available from the first author upon request. Here is the list of reviewed journals in alphabetical order, with number of SEM articles obtained in parentheses: *Canadian Journal of Communication* (0), *Communication Education* (0), *Communication Monographs* (2), *Communication Quarterly* (3), *Communication Reports* (1), *Communication Research* (11), *Communication Research Reports* (0), *Communication*

Studies (0), *Communication Theory* (0), *Critical Studies in Mass (Media) Communication* (0), *European Journal of Communication* (1), *Harvard International Journal of Press/Politics* (0), *Health Communication* (1), *Howard Journal of Communication* (1), *Human Communication Research* (11), *International Journal of Advertising* (2), *International Journal of Public Opinion Research* (5), *Journal of Advertising* (5), *Journal of Advertising Research* (2), *Journal of Applied Communication Research* (5), *Journal of Broadcasting & Electronic Media* (1), *Journal of Communication* (1), *Journal of Communication Inquiry* (0), *Journal of Media Economics* (0), *Journal of Radio Studies* (0), *Journalism & Mass Communication Monographs* (0), *Journalism & Mass Communication Quarterly* (2), *Mass Communication & Society* (1), *Media Psychology* (1), *Media, Culture, & Society* (0), *Newspaper Research Journal* (0), *Political Communication* (1), *Public Opinion Quarterly* (1), *Public Relations Review* (0), *Southern Communication Journal* (0), *Western Journal of Communication* (1), *Women's Studies in Communication* (0).

REFERENCES

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37–47.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 10*, 411–423.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13,* 139–161.

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

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Bielby, W. T., & Hauser, R. M. (1977). Structural equation models. *Annual Review of Sociology, 3*, 137–161.

Blalock, H. M., Jr. (Ed.). (1971). *Causal models in the social sciences.* Chicago: Aldine.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 37–69). Washington, DC: American Sociological Association.

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

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110,* 305–314.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality.* Amsterdam: Sociometric Research Foundation.

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7,* 461–483.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Thusand Oaks, CA: Sage.

Cappella, J. N. (1975). Structural equation modeling: An introduction. In P. M. Monge & J. N. Cappella (Eds.), *Multivariate techniques in human communication research* (pp. 57–110). New York: Academic Press.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 37–59). Thousand Oaks, CA: Sage.

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

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

Davis, D. W. (1997). Nonrandom measurement error and race of interviewer effects among African Americans. *Public Opinion Quarterly, 61,* 183–207.

DeFleur, M. L., & Ball-Rokeach, S. (1989). *Theories of mass communication* (5th ed.). New York: Longman.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Emmers-Sommer, T. M., & Allen, M. (1999). Surveying the effects of media effects: A meta-analytic summary of media effects research in *Human Communication Research*. *Human Communication Research, 25,* 478–497.

Fink, E. L., & Chen, S. (1995). A Galileo analysis of organizational climate. *Human Communication Research, 21,* 494–521.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

Godbold, L. C., & Pfau, M. (2000). Conferring resistance to peer pressure among adolescents: Using inoculation theory to discourage alcohol use. *Communication Research, 27,* 411–437.

Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHD literative stochastic regression imputation and expectation-maximization. *Structural Equation Modeling, 7,* 319–355.

Hackman, M. Z., Ellis, K., Johnson, C. E., & Staley, C. (1999). Self-construal orientation: Validation of an instrument and a study of the relationship to leadership communication style. *Communication Quarterly, 47,* 183–195.

Hale, J. L., Lemieux, R., & Mongeau, P. A. (1995). Cognitive processing of fear-arousing message content. *Communication Research, 22,* 459–475.

Hoyle, R. H. (1991). Evaluating measurement models in clinical research: Covariance structure analysis of latent variable models of self-conception. *Journal of Consulting and Clinical Psychology, 59,* 67–76.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and test of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Comments, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62,* 429–440.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity under parameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351–362.

Hunter, J. E., & Hamilton, M. A. (1995). *PATH: A program in BASICA.* Unpublished manuscript, Michigan State University.

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: Seminar Press.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage.

Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika, 37,* 243–260.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood.* Chicago: National Educational Resources.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide.* Chicago: Scientific Software.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Thousand Oaks, CA: Sage.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Erlbaum.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 16–36). Thousand Oaks, CA: Sage.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

MacCallum, R. C., Roznowski, M., Mar, C. M., & Reith, J. V. (1994). Alternative strategies for cross-validation of covariance structure modeling. *Psychological Methods, 1,* 130–149.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490–504.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Matsueda, R. L., & Bielby, W. T. (1986). Statistical power in covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 120–158). San Francisco: Jossey-Bass.

McCutcheon, A. L., & Nawojczyk, M. (1995). Making the break: Popular sentiment toward legalized abortion among American and Polish catholic laities. *International Journal of Public Opinion Research, 7,* 232–252.

McGuire, W. J. (1986). The myth of massive media impact: Savagings and salvagings. In G. Comstock (Ed.), *Public communication and behavior* (pp. 173–257). New York: Academic Press.

McLeod, J. M., Kosicki, G. M., & Pan, Z. (1996). On understanding and misunderstanding media effects. In J. Curran and M. Gurevitch (Eds.), *Mass media and society* (pp. 235–266). London: Edward Arnold.

McPhee, R. D., & Babrow, A. (1987). Causal modeling in communication research: Use, disuse, and misuse. *Communication Monographs, 54,* 344–366.

McQuail, D. (2000). *Mass communication theory* (4th ed.). London: Sage.

Muliak, S. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Palmgreen, P. (1984). Uses and gratifications: A theoretical perspective. In R. N. Bostrom, (Ed.), *Communication yearbook 8* (pp. 20–55). Beverly Hills, CA: Sage.

Raykov, T., & Marcoulides, G. A. (2000). A method for comparing completely standardized solutions in multiple groups. *Structural Equation Modeling, 7,* 292–308.

Raykov, T., Tomer, A., & Nesselroade, R. J. (1991). Reporting structural equation modeling results in psychology and aging: Some proposed guidelines. *Psychology and Aging, 6,* 499–501.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Newbury Park, CA: Sage.

Scheufele, D. A. (2000). Talk or conversation? Dimensions of interpersonal discussion and their implications on participatory democracy. *Journalism & Mass Communication Quarterly, 77,* 727–743.

Sobel, M. E. (1987). Some new results on indirect effects and their standard errors in covariance structure models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 159–186). Washington, DC: American Sociological Association.

Stephenson, M. T., & Holbert, R. L. (2001). *Analyzing the role of measurement error in three types of structural equation models used in the communication sciences.* Manuscript submitted for publication.

Tanaka, J. S. (1987). How big is big enough?: Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58,* 134–146.

Tanaka, J. S. (1993). Multifacted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Teel, J. E. J., Bearden, W. O., & Sharma, S. (1986). Interpreting LISREL estimates of explained variance in nonrecursive structural equation models. *Journal of Marketing Research, 23,* 164–168.

Thurstone, L. L. (1947). *Multiple-factor analysis.* Chicago: University of Chicago Press.

Wong, C., & Law, K. S. (1999). Testing reciprocal relations by nonrecursive structural equation models using cross-sectional data. *Organizational Research Methods, 2,* 69–87.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.