2017

Covariance-Based Structural Equation Modeling in the Journal of Advertising: Review and Recommendations

Joseph F. Hair

Barry J. Babin

Nina Krey
Rowan University, krey@rowan.edu

Covariance-Based Structural Equation Modeling in the *Journal of Advertising*: Review and Recommendations

Joseph F. Hair, Jr.
University of South Alabama
Mitchell College of Business
Mobile, AL 36688
joefhair@gmail.com

Barry J. Babin
Max P. Watson, Jr. Professor of Business Research
Louisiana Tech University
Ruston, LA 71272
bbabin@latech.edu

Nina Krey
William G. Rohrer College of Business
Rowan University
Glassboro, NJ 08028
krey@rowan.edu

Please cite article as: Joseph F. Hair Jr., Barry J. Babin & Nina Krey (2017): Covariance-Based Structural Equation Modeling in the Journal of Advertising: Review and Recommendations, Journal of Advertising, 46 (1), 163-177.

Contact author: Joseph F. Hair, Jr. (joefhair@gmail.com)

This paper has been accepted for publication in Journal of Advertising. Please note that the final article may include updates to this current version.

ABSTRACT

In this paper, we review applications of covariance-based structural equation modeling (SEM) in the *Journal of Advertising (JA)* starting with the first issue in 1972. We identify 111 articles from the earliest application of SEM in 1983 through 2015, and discuss important methodological issues related to the following aspects: confirmatory factor analysis (CFA), causal modeling, multiple group analysis, reporting, and guidelines for interpretation of results. Moreover, we summarize some issues related to varying terminology associated with different SEM methods. Findings indicate that the use of SEM in the *JA* contributes greatly to conceptual, empirical, and methodological advances in advertising research. The assessment contributes to the literature by offering advertising researchers a summary guide to best practices and reminds of the basics that distinguish the powerful and unique approach involving structural analysis of covariances.

Key Words: Structural equation(s) modeling, SEM, confirmatory factor analysis, CFA, analysis of covariance, theory testing, advertising research

From its very first issue, the *Journal of Advertising* (*JA*) welcomed both innovative analytical approaches as well as reviews of advertising research within its pages. In the first issue, Ginter and Bass (1972) present an innovative approach attempting to establish causality associated with a television advertisement. Other early issues also took the bold approach of publishing review-like comments on empirical work published in the *JA* (Arndt 1972; Largen 1972). In fact, the *JA* explicitly invited such in its very first pages where Sandage (1972, p. 6) states "criticism should be welcomed."

In keeping with this tradition, this paper reviews the use of covariance-based structural equation modeling (SEM) within the extant *JA* volumes. Marketing and consumer research's first applications of modern multivariate statistical procedures, including SEM, date from the 1970s (Aaker and Bagozzi 1979; Darden and Perreault 1975). Given the importance of measurement in advertising research, and the unique contributions of SEM to measure validation, the *JA* was early to publish SEM applications. The first SEM application examines the convergent and discriminant validity of a three-dimensional television ratings measure developed to assess viewer perceptions of advertising relevance, confusion, and entertainment (Lastovicka 1983). Subsequently, more than 100 published *JA* articles apply structural analysis of covariance in one form or another. These articles provide useful content to understand the evolution of SEM in advertising research, as well as the evolution of SEM in business research in general.

This research reviews articles reporting or purporting SEM analyses in the *JA*. We attempt to describe the types of research and the research approaches involved within those articles. In particular, we pay attention to core issues associated with the appropriate use of SEM including practices related to CFA, as the sine qua non of psychometric measurement, causal modeling, and multiple group analyses. This work contributes further to the literature by offering an overview of the technique, illustrations of uses in the advertising literature, and concludes with a summary guide to good practices and insight into other avenues of SEM applications for researchers.

BASIC OVERVIEW

We do not intend detail how to do structural equation modeling, as other reviews or texts provide adequate, although sometimes overlooked, descriptions of the procedure (see Bagozzi and Yi 2012; Byrne 2006; Iacobucci 2010; Shah and Goldstein 2006). Further, the presentation avoids an overly technical or detailed mathematical presentation in keeping with the applied focus of the *JA*. Others provide such presentations (see Bagozzi 1980; Kaplan 2009). We focus on a number of basic issues related to terminology and a fundamental understanding of the technique's capabilities. As the literature evolves, the analytical lexicon can come to be quite confusing and terminology that was clear in the conception of a technique has become cloudy as various researchers adopt the terminology to their own applications. Such an evolution is likely inevitable as a multifaceted tool diffuses ever more widely.

THE STRUCTURE IN STRUCTURAL EQUATION MODELING

The phrase "structural equation model" broadly encompasses an ever increasing family of approaches, statistical, mathematical, and graphical. Similarly, terms like path analysis and factor analysis also display very broad boundaries. In contrast, the term "causal modeling" appears to be used much less often in recent years.

Causal Models

A quick search of the exact term in EBSCOhost reveals six *JA* articles using the term, two appearing since 2000. In the *Journal of Marketing Research*, for example, the term "causal modeling" yields 55 hits, only three since 2000, nine in the 1990s, 39 in the 1980s, and four in the 1970s. The term causal modeling, very familiar to those involved in SEM in the 1980s, clearly has fallen out of favor. Freedman (1991) provides a historical overview of this progression. We would be naïve to deny the fact that authors oftentimes pick terms, and even techniques, as a manner of politically navigating the review process rather than providing the most straight forward description of their research intention (Babin, Griffin, and Hair 2016). Authors may sometimes recognize the limitations in their data, but may have been cautious about stating a causal conclusion even if the original intention was to demonstrate how some change in advertising characteristics brought about a change in performance. Causal conclusions seem central to offering ad managers normative guidance.

The genesis of SEM indeed lies in the desire to draw causal inferences. Pearl (2009; 2010; 2012; 2014) provides a comprehensive review of causality and SEM, and points out that the greater accessibility to statistical techniques coincides with users who may lack a fundamental understanding of key principles. Among these are the relevant assumptions that support causal inferences. Given the inherent limitations that accompany a typical experimental design, particularly with respect to generalizability, researchers long desired to be able to draw causal inferences from nonexperimental data (Blalock 1964; Teas, Wacker, and Hughes 1979). Thus, a tool that could facilitate testing the accuracy of our a priori, hypothetical causal theories indeed represented a major breakthrough for advertising and marketing researchers as it allowed survey-based research to enter the causal domain. In essence, SEM enables us to see how well our preconceived theory of a given set of advertising effects "fits" reality as represented by the observed data. In data terms, SEM allows the theoretical structure of the data (the way the data should look if our explanatory theory is correct) to be directly compared to the actual structure of the observed sample data.

SEM procedures are not the same as typical Ordinary Least Squares (OLS) applications. One key distinction is that in applying SEM, one accounts for non-relationships as well as relationships. In a graphical model, the absence of a connection imposes a constraint that presupposes that these entities are unrelated. In a path diagram, the fact that variables are not connected is just as important as connected variables. One often misunderstood distinction of SEM is the treatment of the error-variance in equations (Pearl 2009). In SEM, error-variance is represented by a latent, exogenous factor. The absence of connections among the error-variance factors represents an independence assumption necessary to establish evidence of non-spuriousness, and thus of causality. That is, the latent factors referred to as error-variance terms capture the effect of all non-measured alternative causes. If the absence of connections, and thus relationships between error-variance factors, is detrimental to fit, causal claims become tenuous. Furthermore, SEM procedures allow us to examine the consequences of a violation of this necessary but insufficient assumption of causality. A relationship between the error-variance factors of a cause and effect means that other common causes likely exist. The model is underspecified. Thus, SEM procedures offer *advantages* in providing evidence of causality (or the lack thereof), and thus the term causal modeling does apply, although causal claims can never be established without logical rationale for causal processes as well.

Structure in SEM

What is structure? If a researcher's theory presupposes the sequence of causal relationships, and equally as important if not more so, the lack of relationships among all the measured variables involved in an analysis, then he or she could presuppose the pattern of relationships in the covariance matrix. As SEM became less commonly known as a causal modeling tool, the question of "what is the structure in the term structural equation modeling?" likewise became less fully known. While today's easy to use software makes tools like SEM accessible to greater and greater numbers of researchers, including those who have not studied multivariate data analysis or SEM in more than a cursory way, users may not comprehend all fundamental concepts. The theoretical structure inevitably coincides with data patterns varying with relationships and/or non-relationships.

AMOS and LISREL, SEM software used nearly exclusively in the *JA*, now enable the user to work directly from raw data.[1] In AMOS, for example, the data to all appearances of the user remain forever in their raw form in SPSS. Thus, the interface can easily lead to the impression that parameters are estimated and answers are derived in the same manner as OLS regression in SPSS. Those using SEM software pre-1990, *fortunately*, did not enjoy that convenient advantage and more clearly understood that covariance provides the foundation of the analysis! Once a user had a theory and corresponding raw data, an initial analysis was needed to extract and store the observed covariance matrix (**S**) of the measured variables in a manner that could be read by SEM software. Therefore, the user was keenly aware that the SEM approach was analysis of *covariance*.

Thus, the *structure* is evidenced in the covariance (or correlation if standardization is used) and its derivatives including factor loading matrices and covariances among unobservables

[1] We limit discussion to the software used by the *JA* authors. AMOS and LISREL are most widely used and are available for use by purchasing a license (as are Mplus and EQS). Many other programs exist now including SEM packages within R, which are free to use. Perhaps the most promising is Lavaan – latent variable analysis.

(Hox and Bechger 1998). In fact, in addition to program execution code, some SEM software provides a matrix entry form with which one could represent the pattern of relationships/non-relationships by specifying proposed patterns in these matrices. The matrices are essential to the matrix algebra equations in structural equation analysis. From a theory-testing standpoint, by far the most critical aspect of an SEM analysis is the model-derived or implied covariance matrix ($\widehat{\boldsymbol{\Sigma}}$). That matrix represents the structure of the data *implied* by the user's theory. When we use the term SEM in the remainder of this article, we refer specifically to analysis of covariance approaches. As the review points out, analysis of covariance SEM is the approach applied most often (nearly universally) in the *JA*. As such, a brief mention of the role of covariance is fundamental (Hair et al. 2017).

The Structure Explains

Researchers interested in cause and effect are motivated by explanation. In fact, theory takes us beyond prediction by offering explanations of not just how much, but why the dependent variable (DVs including endogenous constructs) responds to changes in the independent variables (IVs including exogenous constructs). The analyst must explain if and how one of the IVs, for example K_1, changes systematically with other variables (i.e., K_2). For the theory to be causally complete it must also account for non-relationships; in other words, which variables (observed and latent, including those representing residual variance) do not change in response to others. The program(er) must constrain non-relationships to 0 because they are expected to not exist. Statistically, the structure matches the over-identification that results from constraints corresponding to the theory (Ronkko, McIntosh and Antonakis (2015). Without this *full* accounting of the structure, the explanation is incomplete.

How is a theory tested in a single analysis? By comparing the theorized structure of reality with the observed structure of reality. That is, by comparing the theory-implied covariance structure (matrix) to the observed covariance structure (matrix). The closer the two come to one another, the more accurate is the theory. Thus, the χ^2 statistic, which is relatively simple conceptually, following the functional form,

$$\chi^2 = f(\,|\mathbf{S} - \hat{\boldsymbol{\Sigma}}|\,)$$

becomes the most important outcome in SEM theory testing, and is the most important result to report. As the two matrices become the same, the value tends toward 0. The χ^2 statistic cannot be interpreted without considering parsimony as represented by the model's degrees of freedom (df). While software provides the net degrees of freedom, the model df easily can be computed:

$$df = [(p(p+1)) / 2] - K$$

with p representing the number of measured variables and K the number of free (unconstrained) parameters.

PLACE FIGURE 1 ABOUT HERE

Figure 1 illustrates this process with a basic example. The theory derived model is depicted to show that attitude toward the ad causes changes in attitude toward the brand. $\hat{\boldsymbol{\Sigma}}$, shown in the middle matrix on the right, represents the theorized structure. **S**, the observed data shown in the top right matrix (derived from 475 observations), indeed corresponds fairly closely, but not perfectly. In fact, this result produces a χ^2 of 17.3 with 8 degrees of freedom ($p = .03$). In this case, the actual data structure and theory-derived data structure match pretty closely, or should we say, they 'fit' each other. As alluded to earlier, when the residual covariances exhibit a random pattern, further support of causality is exhibited since such a pattern is evidence against spurious causation. The lack of a connection between the exogenous factors (error-variance for

A_{brand} and A_{ad}) is justified. A key point made by emphasizing the extraction of the covariance matrix is that the moments come from the covariance matrix. Therefore, the degrees of freedom available are determined by the size of the covariance matrix (unique elements only) and not by the sample size. In addition, nothing about the relationships between the indicators (common variance) is lost by the covariances. The covariance matrix contains full information about the data structure and in SEM applications, we end up with poor fit to the extent that our theory does not explain, or match, the observed data. The data structure cannot be separated from SEM analysis and yet be well understood.

Path Analysis

Path analysis, not to be confused with critical path analysis from operations research, is another term widely applied in SEM papers and beyond. The idea is similarly to try to capture the cause and effect paths within a sequence of variables using a cross-sectional analysis. Evans (1978) represents an early "path analysis" approach. The article focuses on the causal paths of a Fishbein model using survey research. The relationships between the beliefs and evaluations components, with attitude toward the ad followed by subjective norms, and ultimately behavioral intentions are estimated using OLS. Thus, what distinguishes this from a traditional OLS regression application? First, cosmetically, the author includes a path diagram showing the flow of effects. In addition, the possibility of indirect effects exists as the multiplicative-product, beliefs*evaluations cause attitude toward the ad, which in turn causes behavioral intent. Second, the path analysis model employs multiple equations, but each equation is independently estimated on the raw data. The main distinguishing characteristic is the emphasis on implied causality and explanation over prediction.

Many *JA* authors employ the term path analysis in their SEM applications. In fact, the use of the term path analysis tends to signify that something other than a covariance-based SEM has been conducted and that typically involves the use of composite factors (most often summated scales), OLS estimates, attenuated estimates (not corrected for measurement error), and sometimes lack of a rigorous psychometric assessment. However, the term path analysis does not coincide with the use of "reduced-form" OLS. A reduced-form approach models the ultimate dependent variable in a single equation that includes only and all exogenous variables as predictors (only a single endogenous factor is involved and; see Cox (2009) for a straight-forward discussion of reduced-form versus structural equation systems). The use of multiple specified equations corresponding to a theory are consistent with the desire to draw causal conclusions.

METHODOLOGY

We set out to identify the articles in *JA* that purport to perform covariance-based structural equation analyses. Articles were identified using key-word searches within the entire *Journal of Advertising* bibliographical record. The key words or phrases include each of the following exact terms:

- Structural equation(s) modeling (modelling)
- SEM
- Confirmatory factor analysis
- CFA
- LISREL
- AMOS
- EQS

- Causal model (modeling)
- Path analysis
- Structural model

Online Appendix shows those articles identified and published between 1983 and 2015, the key word to which it is associated, and a description of the type of analysis involved. The list describes characteristics of the studies relevant to describing how the analysis was actually conducted.

Results

An initial search resulted in a total of 377 articles from the first volume of the *JA* through 2015. After deleting duplicate articles that included more than one search term and articles that only mentioned yet did not apply SEM, the final sample consists of 111 articles matched up to at least one of these key words. Figure 2 plots the frequency of SEM articles' occurrence by year. The plot shows a generally increasing trend with fluctuation from year to year. The greatest frequency of occurrence is 11 SEM articles published in the 2012 calendar year. Given that quite a few of the articles from 1983 to 2015 report multiple SEM applications, the total number of SEM models in the *JA* exceeds 300.

PLACE FIGURE 2 ABOUT HERE

Although advertising research is traditionally both analytical and rigorous, not all advertising researchers are mathematical statisticians. As such, SEM applications grew in proportion to the availability of easy to use software that required neither detailed theoretical knowledge of statistics nor the ability to write program logic. Thus, what SEM software have *JA* authors traditionally employed? Although not all articles report the software applied in the analyses reported, among those that do, more than half (36) indicate LISREL (Jöreskog and

Sörbom 1982). LISREL precipitated the growth of SEM in marketing and has been available almost twice as long as AMOS (and much longer than other software options), so its position as the most often applied software is not surprising. AMOS, with its graphical interface and marketing approach involving a cobranding effort with SPSS, is the second most applied software with 25 applications (see Figure 3). The first AMOS application appears in 2002 and 11 occur since 2010. Among software other than LISREL or AMOS, only EQS shows more than a single application being used in four articles. The distribution of software applications is an indication of the length of availability and the user-friendliness of the software.

With respect to estimation approaches, 90 percent of the applications involve maximum likelihood estimation. Only a handful of the articles report reliance on an SEM with other estimation techniques, such as generalized least squares or distribution-free estimation. The heavy reliance on maximum likelihood is justified based on the relative robustness of the approach (Awang, Afthanorhan, and Asri 2015; Hox and Bechger 1998).

PLACE FIGURE 3 ABOUT HERE

Testing Measurement Theory (Psychometric Validation)

In the early years of SEM applications, researchers had yet to settle on a standardized approach or sequence of steps that would characterize a valid analysis. Chief among debates was how to deal with the measurement theory, sometimes referred to as the auxiliary theory, which specifies how measured (or “manifest”) variables operationalize latent constructs that eventually form the structural theory to be tested (Sajtos and Magyar 2015). The biggest debate boiled down to whether or not the measurement theory should be tested “independently” of the structural theory. After all, the structural theory test is flawed to the extent that measurement is poor. The measurement model indeed serves as an upper bound for the fit of a theoretical model in the sense that a saturated theoretical model, meaning one estimating every possible one-way (i.e., recursive) theoretical relationship, will be the same as that of a standard CFA.

Mathematically the test of correspondence rules linking measured variables to latent constructs and those correspondence rules interconnecting constructs, are not independent. Practically speaking, with good measurement performance, including the lack of evidence of interpretational confounding, the measurement model should be able to stand for examination separately. In fact, should the changes needed to convert a CFA into a recursive theoretical model result in measurement loading estimates changing more than trivially (say > .05), then the analyst should look into the possibility of problems with the measures/data.

An influential article by Anderson and Gerbing (1988) led to a growing consensus that a two-step SEM approach is preferable to a simultaneous test of measurement and structural theory. The two steps constitute, first a test of the theoretical measurement model using confirmatory factor analysis (CFA), and second, only if the CFA displays evidence of construct validity including good fit, a test of a subsequent theoretical or structural model. The measurement theory explains how measured variables represent latent constructs. In the years following publication of this influential article (Anderson and Gerbing 1988), the two-step approach became the standard approach for conducting and presenting SEM results as opposed to the simultaneous approach of testing the model implied measurement and theoretical covariance structures against the observed covariance structure in a single analysis.

CFA fit. CFA assessment is very prominent in the set of relevant *JA* articles appearing after 2000. A total of 87 of the 111 articles report a CFA result. Four articles explicitly are positioned as a scale development where the CFA becomes the key focus of analysis. In particular, the degrees of freedom represent a proxy for complexity because more variables means more degrees of freedom. Extremely simple or complex models can prove problematic. The simplest results presented include models with 0 degrees of freedom. The most complex model contained 1,169 degrees of freedom, while the average is 148. Even though parsimony is desirable, a CFA with 0 degrees of freedom represents a 'just-identified identified' model fitting perfectly by mathematical definition. Thus, a CFA test is not particularly useful if the fit is predetermined. The notion of identification may not be well understood by all users and reviewers as at least three CFAs are reported with 0 degrees of freedom (see for example, Sirgy et al. 1998). Likewise, models with few degrees of freedom are very simple and by that fact alone should provide relatively good fit. Two additional applications report 5 degrees of freedom or fewer (Noguti and Russell 2014). While on a rare occasion one might find value in trying to validate a single 4-item dimension alone (perhaps if the paper reports the development of such a scale), one should expect an insignificant p-value if there is any case at all to be made for good fit. Simple models should be held to the strictest standard.

How did the CFA models perform? Prime facie evidence of good fit results from an insignificant χ^2 statistic, thus signaling that the implied covariance matrix, computed from the theoretical model representation, is not significantly different from the observed covariance matrix. Twenty-five (12 percent) of the reported χ^2 statistic values, excluding those with 0 values, are insignificant based on a p-value greater than 0.05. As a frame of reference for future users, the unweighted average is 561 with 147 degrees of freedom.

A host of various other fit indices developed over time, motivated in great part by the fact that the χ^2 test, like other parametrical statistical tests, is susceptible to power effects and quickly becomes significant with large samples. Early on, little consensus existed over which of these indices was most appropriate to report. Over time, the Comparative Fit Index (CFI) (Bentler

1990) and the Root Mean Squared Error of Approximation (RMSEA) became the most commonly referred to goodness and badness of fit indices, respectively. Hair et al. (2010) provide a discussion of the evolution of these statistics and guidelines for assessing values for CFI and RMSEA for models with varying degrees of complexity and varying sample sizes. The key idea of their analysis is that there is no one size fits all cutoff value for these statistics, meaning it is impossible to apply a single standard to all models and research situations. Complicated models, consisting of more variables and constructs, with large sample sizes could fit well with lower CFI values than simple models with small samples. The lower value for CFI associated with good fit for the most complicated models is approximately 0.92. In the articles published/featured in the *JA* reporting only CFA results, the CFIs range from 0.75 to 1.0, with an average CFI of 0.949, and a RMSEA ranging from 0 to 0.22, and an average of 0.081. Thus, if extreme values are excluded, most of the CFA models are within the guidelines of good fit. Further, researchers should report the initial CFA results as well as the results subsequent to substantial modification of the theoretical measurement model. In addition, at times authors may report competing theories, such as comparing two-factor to one-factor solutions as a test of unidimensionality. These tests account for some of the low goodness of fit values reported in the *JA* (e.g., Latour and Rotfeld 1997). Additionally, the χ^2 value Like for other parametric statistics, the p-value itself is of limited value, particularly in situations with high statistical power. But, reporting both the χ^2 and df should be an essential part of any report.

Measurement Theory. Valid measurement is a nonnegotiable characteristic of good research. Historically, measurement validity was based on face and content validity that resulted from a qualitative assessment. With the adoption of SEM, measurement validity is also assessed quantitatively. Thus, good psychometric measurement is characterized by evidence of fit validity as demonstrated empirically by favorable CFA results. CFA fit validity is a necessary (but insufficient) condition for overall validity, so long as the measurement theory allows each indicator to load on only one factor and constrains all error-covariance to 0 (i.e., an appropriate congeneric representation) (Babin and Zikmund, 2016). This means the hypothesized measurement structure (leading to the implied covariance matrix) closely mirrors the observed covariance matrix. Another measurement theory criterion is convergent validity, meaning the items representing a construct correspond with each other to represent a unidimensional factor. Beyond fit validity and convergent validity, discriminant validity should be present, so that each measured variable corresponds to only a single construct and the constructs that make up a model each represent a unique entity. Other forms of validity such as concurrent and predictive are possible but much less often associated with SEM.

CFA, properly applied, is a statistical tool uniquely qualified to provide empirical evidence of validity for any set of latent constructs. Some of the *JA* articles do not indicate a properly applied CFA model. Sometimes, limitations of the research prevent a full examination of the validity of measures. At other times, a CFA can be conducted in a way that does not allow a full test of measurement validity, meaning one or more aspects crucial to validation are impossible to detect. One such practice is the use of partial CFA. Partial CFA involves measurement models containing only a portion of the measurement model involved in a study; sometimes testing model constructs in a model individually rather than as part of an integrated model. Thus, a five-construct model will include five separate CFAs. While this approach may provide evidence of convergent validity within each factor, it is not possible to examine discriminant validity. A partial CFA completely masks any lack of fit that would be produced if the indicators of one construct covary strongly with the indicators of another construct.

Similarly, a strong correlation suggesting a lack of discriminant validity between constructs also would go undetected.

Based on the descriptions of the CFA models reported in the *JA* articles, as many as 30 report some form of a partial CFA. As an example, one article assesses multiple constructs but reports a CFA for only one construct – ad-evoked emotions (Zhao, Muehling, and Kareklas 2014). The ad-evoked emotions construct includes item indicators, such as happy and pleased. The authors also measure brand attitudes with items such as “good” and “favorable,” but do not include attitude and emotion together in a CFA. In sum, all of the items and constructs should be included in an overall CFA to address any concerns about construct validity, whether convergent or discriminant.[2] Assessment of differences in fit based on additions of cross-loadings or consolidation of constructs could be helpful in such an assessment, but may create other problems (O'Rourke and Hatcher 2013). If a model with cross-loadings (an item reflecting more than one construct) fits better than one without, the results provide evidence of a lack of construct validity.

In a small number of cases, CFAs were reported using composite scales. This approach also masks problems with individual items, such as strong residual covariance that would diminish fit. Thus, the general rule for CFA is to include all latent constructs and variables involved in the theoretical model in a single CFA. Note that if an appropriate single item variable is in the model (e.g., sales), it may not be included in the CFA. Moreover, in rare cases a parceling approach could be adopted. Parceling involves taking composite subsets of item

[2] Convergent validity, the extent to which multiple measures converge on a consistent meaning, and discriminant validity, the extent to which a measure is unique and not confounded by another, are both necessary elements of the broader concept of construct validity, the extent to which a measure truly represents a construct.

indicators and using them in place of the individual item indicators for a construct. For example, a construct measured by 50 items might be modeled with ten indicators that are each composites of five measured variables. In no case in the *JA* articles was a scale long enough to call for item parceling and rarely in advertising or marketing research would one find such a long scale.

Identification. Another issue that limits a model's ability to fully test measurement validity involves statistical identification (Blalock 1964). A full description of statistical identification is well beyond the scope of this article, but the crux of the matter involves under-identification, an inability to mathematically derive a solution due to insufficiencies such as a lack of information. The most common statistical violation occurs when a researcher, usually without full awareness, proposes a theory that requires more parameters to be estimated than the number of unique moments used as input to the model (i.e., the number of unique elements of a covariance matrix). For example, if one tries to estimate a single factor model indicated by two measured variables, no standard CFA solution is possible because mathematically, the model is under-identified (the model requires four parameters to be estimated while the covariance matrix providing input to the model contains only three unique moments). If the researcher combines the two-item construct into a model consisting of several other constructs, each measured by several items, the overall model may become identified making a solution possible. However, statistical identification of the two-item construct remains problematic. Thus, problems with model convergence or unstable solutions are often attributable to an identification problem. More than 25 of the *JA* articles include measures with less than three-items as indicators. A comprehensive discussion of identification can be found in Mardia, Kent, and Bibby (1980).

Even single-item measures can be included in an SEM model. But the loading parameter and associated error-variance parameter should be specified rather than estimated. In some cases, authors indicate fixing parameters for single-item measures. A fixed parameter is one that is constrained to some predetermined value. If one assumes perfect measurement, the loading for a single-item can be fixed to 1 and the error variance fixed to 0. But the fixing of parameters is not typically described for two-item measures. As such, the risks of an unstable solution for models with two-item factors remains high, and the results are not as dependable as a model that would be fully or over identified overall and within each theoretical construct. For this reason, aside from any theoretical arguments about the number of items sufficient to measure any given construct, a measure with a minimum of three measured indicators is greatly preferable and will minimize problems as researchers perform analyses.

One article illustrates problems related to a lack of statistical identification (Leigh, Zinkhan, and Swaminathan 2006). The authors report difficulties such as a "small, nonsignificant negative error variance" (p. 115), otherwise known as a Heywood case. Equivalently, a negative error variance means a quirky result indicating more than 100 percent explained variance. Additionally, subsequent models report a path estimate (presumably standardized) of 0.99 between cognitive and recall constructs. The observed correlations for the measured variables corresponding to this relationship are between 0.5 and 0.75. Thus, the 0.99, signaling that the two constructs are synonymous, is likely an unreliable estimate due to the lack of identification. Similar to single-item factors, if a two-item factor is included, additional parameters must be fixed (fixed means constrained) rather than freed. For instance, both factor loading estimates can be fixed to some like value (e.g., 0.70) rather than estimated.

Thus, although most of the articles using CFA exhibit good practices, researchers can learn from awareness of some of the more questionable approaches. One trend discovered in the review that is difficult to quantify is the tendency of researchers to delete items in the process of doing a "confirmatory" factor analysis. In some cases, the authors do not clearly state how many items were discarded in the process of trying to confirm a factor structure. Such information is very important for others who may wish to use the same scales or replicate the analysis. In the discussion, we'll follow up with thoughts on when modifications are numerous enough that the analysis is no longer "confirmatory."[3]

Theoretical Models

Reviewing SEM applications published in the *JA*, at least 60 of the 300 plus applications across the 111 articles report a theoretical (between constructs) model fit without a corresponding CFA result. The bulk of those not reporting a CFA are published in issues before 2000. Thus, the influence of the two-step approach is seen in the later years. Moreover, there are other potentially questionable approaches or unclear reporting. In some cases, the number of items reported for a measure does not correspond to the degrees of freedom reported in a model. At times, composite indicators are used to represent constructs actually measured by a battery of several items (i.e., Muehling, Lazniack, and Stoltman 1991). While such an approach produces very parsimonious tests of theoretical models (2 degrees of freedom in this case), the drawback, as with partial CFAs, is the possibility that problems with item validity are masked by the composite. Issues with discriminant validity, such as when a single item relates highly to two constructs rather than one, not only are hidden in this approach but also produce a higher

[3] We decided not to detail the distinction between reflective and formative indicators. As is expected, and likely is appropriate given the perceptual nature of most of the research, the vast majority of measures involve reflective indicators. Fewer than five studies state some type of formative measure. However, one misnomer is that SEM is not appropriate for formative indicators. Formative indicator models present problems with statistical identification unless formulated in a manner as to avoid these problems. Thus, caution is advised to make sure over-identifying assumptions are in place For a more comprehensive explanation, we refer you to a source such as MacKenzie, Podsakoff, and Podsakoff (2011).

structural parameter coefficient than would be observed otherwise. Thus, a CFA should be encouraged whenever possible.

Among all models of causal relationships across constructs, how well do the *JA* authors' theoretical models tend to fit? The averages provide some benchmark values. The unweighted average χ^2 statistic is 219.5 with 78 degrees of freedom. Both are less than the average values for the reported CFA models above. The average reported CFI is 0.944, and ranges from 0.66 to 1.0. The average reported RMSEA is 0.087, with a range from 0 to 0.5. Based on these results, the majority of model results appear to fall within the rules of thumb for good fit, and drawn from peer review sources. The reason theoretical models, on average, report fewer degrees of freedom than CFA models is that in many cases composite indicators are used in the theoretical model, with or without a CFA including individual item indicators.

Early in the evolution of SEM, the use of composites (combinations of indicators, most often average summated scores) in theoretical models was condoned if an acceptable CFA model fit (with individual item indicators) was first presented. One motivation for such practice is the pursuit of an insignificant χ^2 statistic, which is much more likely in a model with relatively few degrees of freedom. In the extreme, a model that could consist of dozens of items is reduced to a saturated theoretical model with each construct represented by composites resulting in 0 degrees of freedom and "perfect" fit (Henthorne, Latour, and Natarajan 1993). Thus, the switch from individual items in the CFA to composites in the theoretical model is essentially a cosmetic change that produces a more enticing appearance based on over attention to p-values.

One aspect that is lost in such an approach, and often overlooked otherwise, is the value of the CFA model in assessing theoretical fit. All of the *JA* applications involve recursive models, meaning models in which the flow of causation is only in one direction (no reciprocal causation exists). In a recursive model, the CFA provides an upper bound (i.e., the best) on the fit of a subsequent theoretical model. Therefore, a theoretical model can fit no better than the corresponding CFA model. Consequently, the CFA fit provides a basis with which to judge the subsequent fit of the theoretical model. The closer the two become, the better the fit of the theoretical model. Given that the CFI attempts to lessen the effects of sample size and model complexity on fit, it provides a basis for a potentially useful index for assessing theoretical fit following a CFA comprised of the same indicators. The following theoretical fit index (TFI) can be used for that purpose:

$$TFI = \left(\frac{CFI_{CFA} - CFI_{TM}}{CFI_{CFA}}\right)$$

When evaluating the TFI, relatively small values indicate a better theoretical model fit. For example, given a CFA with CFI of 0.99, a subsequent theoretical structural model (TM) of 0.97 yields a TFI of 0.02, approximately a two percent drop in fit. In contrast, if the theoretical model fit is a CFI of 0.92, the TFI would be 0.07, representing approximately a seven percent drop in fit. The TFI works as a badness of fit indicator because a higher value means a relatively worse fit. Thus, the former model would provide a much better theoretical fit. Here again, one should always take parsimony into account. A theoretical structural model with only 1 degree of freedom difference from the CFA model is very nearly a reproduction of the measurement model. As the number of degrees of freedom difference increases, the theoretical model is more parsimonious and larger differences in fit are to be expected. Thus, an Adjusted TFI is proposed, which includes an adjustment for parsimony:

$$ATFI = \left(\frac{CFI_{CFA} - CFI_{TM}}{CFI_{CFA}}\right) X \frac{DF_{TM}}{DF_{CFA}}$$

The use of TFI or ATFI applies to models following the two-step SEM approach, and if it is to be used to demonstrate that the tested theory fits, it is predicated on a CFA with good fit. When appropriate, researchers should consider assessing and reporting the ATFI as an indicator of theoretical model fit. The ATFI always should be interpreted in light of the overall CFA fit. A good structural model ATFI with a poor measurement model remains tenuous.

Conceptually, the breakthrough represented by the goodness of fit test cannot be overstated. No longer were researchers relegated to assessing relationships one at a time or for a few variables predicting a single dependent variable in a multiple regression model. Even then, the analysis is restricted to overlapping variance with inherent limitations in explanatory power. The χ^2 goodness of fit test is the gauge by which one's theory is assessed. Now, one's theory can be represented by a theoretical structure (covariance matrix) and compared against an actual structure (covariance matrix) obtained from sample data. Since the deductive structure matches the data derived in the real world, the theory becomes validated.

Moreover, the comparison of model fits using the χ^2 difference statistic can be extremely useful in assessing which of multiple, plausible, theoretical processes most closely matches reality. For instance, a parameter can be fixed to be equal across two cultural groups. Should the constraint worsen fit, as evidenced by a change in the χ^2, the result presents evidence of moderation by culture.

Multigroup Analyses

Multigroup SEM analysis is particularly relevant to advertising research. Multigroup SEM models are aimed at simultaneously trying to reproduce the covariance matrix for each group examined. Research reported in the *JA* often involves cross-cultural comparisons with samples from two or more countries (cultural regions) or a moderating group effect. Multigroup

SEM is an ideal tool to report results in either situation. Over the years, the multigroup approach has become useful and relatively standardized. Babin, Borges, and James (2016) provide a recent review of the methodology and demonstrate the versatility of multigroup SEM.

Not surprisingly then, *JA* authors employ multigroup SEM analysis and 14 articles refer to it in their methods in some way. Andrews (1989) represents an early application of a multigroup SEM analysis examining potential invariance in the factor structure of a seven-item scale of attitudes toward advertising. Thus, its publication is another instance demonstrating *JA*'s contribution to innovation in advertising research. The article predates the adoption of standard methodology and terminology. One key term is metric invariance. Metric invariance refers to the fact that both a factor loading structure and relative size of loadings does not change (is not significantly different) from group to group. Although the authors' intent appears to have been to examine whether the factor structure remained constant across groups, the number of degrees of freedom reported suggest that some other constraints were imposed. Kim, Baek, and Choi (2012) also report a two-group factor analysis testing metric invariance. The results appear to be reported clearly (p. 85, 86) although with six factors and 16 items, under-identified factors are involved in the analysis. The authors then switch to a "path analysis" using multiple regression to examine structural relationships across groups, attributing the decision to a small sample size. Perhaps other factors played into that decision since the smallest group size is 148, seemingly sufficient for the model to be examined (MacCallum et al. 2001). Other authors point to difficulties with model convergence in their attempts to use a multigroup analysis (Garretson and Niedrich 2004). As a consequence, they also applied path analysis to test for moderation. Authors using multigroup analysis need to pay particular care to the descriptions of the goals of an analysis and their results to reduce ambiguity in meaning. In addition, multiple group analyses, although not terribly difficult to execute with software, can be intricate and benefit from experience with the approach.

DISCUSSION

Structural equation modeling is a very useful research tool in the advertising literature based on trends outlined in the review. In the past, many social scientists considered SEM to be quite a complex statistical method, requiring considerable investment in time and effort to master. Most social science disciplines took years to disseminate the concepts and absorb the knowledge, and to some extent still struggle with advanced statistics. Although readily accessible software, in a sense, brings advanced statistical techniques to the masses, getting the software to produce results does not constitute understanding of the technique. Thus, it is not surprising to see similar challenges faced by advertising.

The attractiveness of SEM for advertising scholars and practitioners can be attributed to the method being an excellent tool to examine and test advertising theories. Those theories include measurement theories where *JA* authors should present research using CFA to provide other researchers with better psychometric scales or to corroborate existing scales for measures relevant in explaining why advertising works or does not work. Moreover, it has great potential to facilitate better understanding of which theories are appropriate for a particular situation, how to improve pricing of advertising, how to evaluate advertising strategies, explaining customer responses to varied stimuli, and so forth. Authors and reviewers should be more open to comparing models based on truly competing theories, where the relative fit can be used as evidence of which explanation best matches reality. In addition, the ability to replicate results from prior studies is *greatly* facilitated by the ability to compare model fit across samples as illustrated in a multigroup analysis. The extent to which prior studies can be reproduced is important to the credibility of science (Hubbard 2016). SEM provides a tool well suited to assist in this important endeavor and more efforts to compare results of theory tests, including replication using SEM, presents an important avenue for future research.

Our review of SEM applications in the *Journal of Advertising* indicates clearly that the use of SEM in advertising contributes greatly to conceptual, empirical, and methodological advances in advertising. Indeed, the trend toward improving knowledge and overcoming weaknesses in advertising research is evident. Like any academic journal, we also can learn from critically reviewing papers that are published. In short, it appears SEM's theoretical methodology, like with other statistical techniques, is sometimes not well understood by its users. A potential reason for unfamiliarity with SEM principles might be that multivariate data analysis textbooks did not traditionally discuss SEM at all, or present the coverage of the topic in an understandable manner given many advertising and marketing scholars' limited mathematical training. Universities should at minimum ensure that all graduate research students learn to understand and appreciate the strengths and limitations of SEM.

Statistical approaches are like tools in a mechanic's toolbox. Although in any situation there may be more than one tool that can at least "sort of" get the job done, the good mechanic knows the right tool for the right problem. Likewise, advertising researchers must understand the initial goal of data analysis is to select an appropriate statistical method. Choose the wrong tool or apply it poorly and at best, one makes life difficult and at worst, the results do not paint an accurate picture. Also, we are wise to remember the words of Clint Eastwood stating: "a man's (researcher) got to know his limitations." Researchers should use the tools they understand, or collaborate with a coauthor that truly understands the desired approach.

Many of these issues were discussed in our review, such as removing too many indicators as a means of achieving fit, lack of congruence based on degrees of freedom, and limitations of understanding constraints in multigroup comparisons. These topics have been researched in the methodological literature, and yet they are being overlooked or just not reported. Advertising researchers are well advised to more strongly consider the methodological foundations of the SEM method and complementary analysis techniques.

To emphasize practical issues and decisions associated with the application of SEM, and to highlight ones that advertising researchers should consider more closely, we suggest the following framework for decision-making when using SEM. Specifically, SEM applications should follow a step-by-step process, as shown below:

1. Model Specification – to match a theory
2. Model Identification – to identify adequate data
3. Model Estimation – to provide parameter estimates and $\hat{\boldsymbol{\Sigma}}$, as a barometer for the theory
4. Model Evaluation – to assess fit and other aspects of validity
5. Model Re-specification – to compare theoretical explanations, further examine conditions of causality, explore post-hoc results or improve model fit (only to the extent where changes are minor and do not fundamentally change meaning; non-minor modifications result in a shift toward developmental or post-hoc results rather than theory testing). Cross-validation using new data when possible.
6. Model Reporting – to draw appropriate conclusions

Each step includes multiple decisions that have implications for subsequent steps. Incorrect decisions in an earlier step can create problems in subsequent steps. Indeed, incorrect decisions in earlier steps may result in invalid model results. Other sources provide a more complete description of these assumptions (see Kaplan 2009; Kline 2015; MacKenzie, Podsakoff, and Podsakoff 2011; McDonald and Moon-Ho 2002).

SEM offers many beneficial properties for testing measurement theory and subsequently, theories comprised of more than individual relationships (e.g., Babin, Boles, and Hair 2008; Bagozzi and Yi 1988; Chin, Peterson, and Brown 2008). Indeed, SEM is not the tool for the researcher who is exploring data. SEM also is not the tool for someone interested only in mere prediction. Having found through CFA that the measures are lacking, SEM is not a good tool for research plagued with poor measures. In Table 1, we provide basic suggestions for good practice including various rules of thumb motivated largely by issues discovered in our review. Others are more broadly applicable. By following these recommendations, we wish to reduce the danger created by easy to use software that researchers without proper foundation are getting results from but not understanding how they came to be.

PLACE TABLE 1 ABOUT HERE

We strongly recommend that authors, reviewers, and editors be familiar with and observe these guidelines. High quality peer-reviewed journals should more strongly emphasize the importance of adhering to these guidelines. Moreover, making all information available, including the data used in the analysis, will facilitate the replication of statistical analyses (e.g., in Web Appendices). Progressing towards the highest possible level of transparency will substantially improve the way research is conducted, and give a free rein to accelerated development paths for methodological issues increasingly encountered in social sciences research.

Conclusion

This article contributes to the advertising and marketing literature through improved implementation of SEM procedures in the advertising literature and beyond. Although we do not intend to settle terminology issues, we also wish to call attention to some potential boundaries in terms. In this article we have used, perhaps in what some may call a past practice, the term SEM to refer to covariance-based SEM. In fact, in previous eras the term Analysis of Covariance Structures was synonymous with CFA and SEM procedures. Today, other linear modeling, graphical, non-parametric, and probability-based approaches fall under the SEM rubric. In addition, the term path analysis (as opposed to SEM) is more commonly used to refer to models with multiple-item indicators based on composites and not common factors where the aim is to make a causal inference. Confirmatory factor analysis applies to an approach that tests measurement theory by examining the fit of the proposed theoretical measured model and not to a set of composites. Causal modeling is a term that can apply to SEM, but only if conditions beyond those necessary to generate estimates are met. In particular, counterfactual conditions implying the lack of relationships are a key in having SEM allowing causal inferences (Pearl, 2009). This type of counterfactual account is imperative to understand the nature of causality in structural models (Pearl 2014). Indeed, connections that do not exist are critically important and represent assumptions that are generally only implied. In fact, the tradition of stating models piecewise with individual hypotheses for each path contributes to a misunderstanding of this fundamental point by drawing attention to the connections and away from the assumptions that affect the theory's interpretation. To the extent that advertising benefits from theories of cause and effect, SEM remains the appropriate tool.

This article emphasizes the fit concept as a key distinguishing factor of covariance-based SEM and links it to the covariance structure. We also introduce the notion of a Theoretical Fit

Index (TFI), and its parsimony adjusted variation the ATFI, as an alternative and potentially improved way of assessing fit of the theoretical model subsequent to a CFA. Researchers are encouraged to use this index. SEM is an explanatory tool. Recall philosophically that explanation begets prediction but not vice-versa (Hunt 2010). The covariance-based SEM user carries the attempts to account for all the information in the covariance matrix. Any deficiency in how one variable corresponds (or does not correspond) to another, costs the theory in terms of fit. Thus, SEM is a tool used when a researcher wishes to test an explanatory theory.

In addition, we review SEM applications in the *JA* as a motivation for suggestions for improvement. We would be remiss to not to point out that there can be too much emphasis on a benchmark of "good models" or "good fit" for publication. If a theory is strong, knowing that it does not fit is equally as important as knowing that it does fit. However, the academic community possesses a strong publication bias that suggests reviewers only consider good results to be those supported by statistical significance, or in the case of SEM, good fit. This kind of thinking, although perhaps also born in love of a given theory, leads to what some call advocacy in research such that researchers are only willing to report "good" results that support the preconceived theory (Babin, Griffin, and Hair 2016; Woodside 2016). If the theory is relevant and the research is done well, then test results should not determine publication. Only in this way do we openly encourage truthful reporting of results. In addition, research should be replicable. SEM provides a great way to examine replication because a study done with a second sample, or hold out sample, can be directly compared and even tested for cross-validity using a multigroup approach and analysis of fit. We must also embrace surprising results that come out of our models and not force fit SEM as a statistical tool for every application. Knowing when not to use a tool is equally as important as knowing when to use that tool! Further, SEM will also play a role in big data analysis as we recognize ways to measure latent phenomena residing in sets of variables recorded through online behaviors. SEM has become an essential tool in the advertising researcher's statistical toolbox, and it will continue to be important to advertising in the future.

REFERENCES

Aaker, David A., and Richard P. Bagozzi (1979), "Unobservable Variables in Structural Equation Models with an Application in Industrial Selling," *Journal of Marketing Research*, 16 (2), 147-58.

Anderson, James C., and David W. Gerbing (1988), “Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach,” *Psychological Bulletin*, 103 (3), 411–23.

Andrews, J. Craig (1989), "The Dimensionality of Beliefs Toward Advertising in General," *Journal of Advertising,* 18 (1), 26-35.

Arndt, Johan (1972), "Intrafamilial Homogeneity for Perceived Risk and Opinion Leadership," *Journal of Advertising*, 1 (1), 40-47.

Awang, Zainudin, Asyraf Afthanorhan, and M. Asri (2015), “Parametric and Non Parametric Approach in Structural Equation Modeling (SEM): The Application of Bootstrapping,” *Modern Applied Science*, 9 (9), 58-67.

Babin, Barry J., James Boles, and Joseph F. Hair (2008), “Publishing Research in Marketing Using Structural Equation Modeling,” *Journal of Marketing Theory and Practice*, 16 (4), 279-85.

______, Adilson Borges, and Kevin James (2016), “The Role of Retail Price Image in a Multi-Country Context: France and the USA,” *Journal of Business Research*, 69 (3), 1074-81.

______, Mitch Griffin, and Joseph F. Hair (2016), "Heresies and Sacred Cows in Scholarly Marketing Publications," *Journal of Business Research*. In press. http://dx.doi.org/10.1016/j.jbusres.2015.12.001

______, and William Zikmund (2016), *Exploring Marketing Research*, Mason, OH: Cengage.

Bagozzi, Richard P. (1980), *Causal Models in Marketing*, Hoboken, NJ: John Wiley & Sons, Inc.

Bagozzi, Richard P., and Youjae Yi (1988), “On the Evaluation of Structural Equation Models,” *Journal of the Academy of Marketing Science*, 16 (1), 74-94.

______, and Youjae Yi (2012), "Specification, Evaluation, and Interpretation of Structural Equation Models," *Journal of the Academy of Marketing Science*, 40 (1), 8-34.

Bentler, Peter M. (1990), “Comparative Fit Indexes in Structural Models," *Psychological Bulletin,* 107 (2), 238-46.

Blalock, Hubert M. (1964), *Causal Inferences in Nonexperimental Research,* Chapel Hill, NC: University of North Carolina Press.

Chin, Wynne W., Robert A. Peterson, and Steven P. Brown (2008), "Structural Equation Modeling in Marketing: Some Practical Reminders," *Journal of Marketing Theory and Practice*, 16 (4), 287-98.

Cox, L.A. (2009), "Why Reduced-Form Regression Models of Health Effects Versus Exposures Should Not Replace QRA: Livestock Production and Infant Mortality as an Example," *Risk Analysis*, 29 (12), 1664-71.

Darden, William R., and William D. Perreault Jr. (1975) "A Multivariate Analysis of Media Exposure and Vacation Behavior with Life Style Covariates," *Journal of Consumer Research,* 2 (2), 93-103.

Evans, Richard H. (1978), "Planning Public Service Advertising Messages: An Application of the Fishbein Model and Path Analysis," *Journal of Advertising*, 7 (1), 28-34.

Freedman, D. A. (1991), "Statistical Models and Shoe Leather, with Discussion," *Sociological Methodology*, 21(2), 291-313.

Garretson, Judith A., and Ronald W. Niedrich (2004), "Spokes-Characters: Creating Character Trust and Positive Brand Attitudes," *Journal of Advertising*, 33 (2), 25-36.

Ginter, James L., and Frank M. Bass (1972), "An Experimental Study of Attitude Change, Advertising, and Usage in New Product Introduction." *Journal of Advertising,* 1 (1), 33-39.

Green, Elliott (2016), "What are the Most Cited Publications in the Social Sciences," *The London School of Economics and Political Science*, May 12, http://blogs.lse.ac.uk/impactofsocialsciences/2016/05/12/what-are-the-most-cited-publications-in-the-social-sciences-according-to-google-scholar/.

Hair, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson (2010), *Multivariate Data Analysis*, Upper Saddle River, NJ: Prentice Hall.

______, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt (2017), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd edition, Thousand Oaks, CA: SAGE Publications, Inc.

Henthorne, Tony L., Michael S. LaTour, and Rajan Nataraajan (1993), "Fear Appeals in Print Advertising: An Analysis of Arousal and Ad Response," *Journal of Advertising*, 22 (2), 59-69.

Hox, J. J., and Bechger, T. M. (1998), “An Introduction to Structural Equation Modeling,” *Family Science Review*, 11, 354-73.

Hunt, Shelby D. (2010), *Marketing Theory: Foundations, Controversy, Strategy, Resource-Advantage Theory,* Armonk, NY: ME Sharpe.

Iacobucci, Dawn (2010), "Structural Equation Modeling: Fit Indices, Sample Size, and Advanced Topics," *Journal of Consumer Psychology*, 20, 90-98.

Jöreskog, Karl G., and Dag Sörbom (1982), "Recent Developments in Structural Equation Modeling," *Journal of Marketing Research*, 19 (4), 404-16.

Kaplan, David (2009), *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: SAGE Publications, Inc.

Kline, Rex B. (2015), *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Kim, Jooyoung, Youngshim Baek, and Yang Ho Choi (2012), "The Structural Effects of Metaphor-Elicited Cognitive and Affective Elaboration Levels on Attitude Toward the Ad," *Journal of Advertising,* 41 (2), 77-96.

Largen, Robert G. (1972), "CRITIQUE," *Journal of Advertising*, 1 (1), 45-46.

Lastovicka, John L. (1983), "Convergent and Discriminant Validity of Television Commercial Rating Scales," *Journal of Advertising*, 12 (2), 14-52.

LaTour, Michael S., and Herbert J. Rotfeld (1997), "There are Threats and (Maybe) Fear-Caused Arousal: Theory and Confusions of Appeals to Fear and Fear Arousal Itself," *Journal of Advertising*, 26 (3), 45-59.

MacCallum, R.C., K.F. Widaman, K.J. Preacher and S. Hong (2001), “Sample Size in Factor Analysis: The Rodel of Model Error,” *Multivariate Behavioral Research*, 36 (4), 611-37.

MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff (2011), “Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques,” *MIS Quarterly*, 35 (2), 293-334.

Mardia, Kanti V., John T. Kent, and John M. Bibby (1980), *Multivariate Analysis*, London, UK: Academic Press.

McDonald, Roderick P., and Ringo Ho Moon-Ho (2002), “Principles and Practice in Reporting Structural Equation Analyses,” *Psychological Methods*, 7 (1), 64-82.

Muehlin, D., Russell N. Laczniak, and Jeffrey J. Stoltman (1991), "The Moderating Effects of Ad Message Involvement: A Reassessment," *Journal of Advertising*, 20 (2), 29-38.

Noguti, Valeria, and Cristel Antonia Russell (2014), "Normative Influences on Product Placement Effects: Alcohol Brands in Television Series and the Influence of Presumed Influence," *Journal of Advertising*, 43 (1), 46-62.

O'Rourke, Norm, and Larry Hatcher (2013), *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling*, Cary, NC: SAS Institute, Inc.

Pearl, Judea (2009), “Causal Inference in Statistics: An Overview,” *Statistical Surveys*, 3, 96-146.

_______ (2010), “The Foundations of Causal Inference,” *Sociological Methodology*, 40 (1), 75-149.

______ (2012), “The Causal Foundations of Structural Equation Modeling,” in *Handbook of Structural Equation Modeling*, R. H. Hoyle, ed., New York, NY: Guilford, 1-37.

______ (2014), “Interpretation and Identification of Causal Mediation,” *Psychological Methods*, 19 (4), 459-81.

Ronkko, M., C.N. McIntosh and J. Antonakis (2015), “On the Adoption of Partial Least Squares in Psychological Research: Caveat Emptor,” Personality and Individual Differences,” Personality and Individual Differences, 87, 76-84.

Sajtos, Laszlo, and Bertalan Magyar (2015), "Auxiliary Theories as Translation Mechanisms for Measurement Model Specification," J*ournal of Business Research*. In press.

Sandage, Charles H. (1972), "Some Institutional Aspects of Advertising," *Journal of Advertising*, 1 (1), 6-9.

Shah, Rachna, and Susan Meyer Goldstein (2006), “Use of Structural Equation Modeling in Operations Management Research: Looking Back and Forward,” *Journal of Operations Research*, 24 (2), 148-69.

Sirgy, M. Joseph, Lee Dong-Jin, Rustan Kosenko, H. Lee Meadow, Don Rahtz, Muris Cicic, Jin Guang Xi, Duygun Yarsuvat, David L. Blenkhorn, and Newell Wright (1998), "Does Television Viewership Play a Role in the Perception of Quality of Life?" *Journal of Advertising*, 27 (1), 125-42.

Teas, R. Kenneth, John G. Wacker, and R. Eugene Hughes (1979), "A Path Analysis of Causes and Consequences of Salespeople's Perceptions of Role Clarity," *Journal of Marketing Research*, 16 (3), 355-69.

Woodside, Arch, G. (2016), “The Good Practices Manifesto: Overcoming Bad Practices Pervasive in Current Research in Business,” *Journal of Business Research*, 69 (2), 365-81.

Zhao, Guangzhi, Darrel D. Muehling, and Ioannis Kareklas (2014), "Remembering the Good Old Days: The Moderating Role of Consumer Affective State on the Effectiveness of Nostalgic Advertising," *Journal of Advertising*, 43 (3), 244-55.

TABLE 1

Some Suggestions for Good Practice

Issue	Comment	Suggestion/Rules of Thumb
Setup and Reporting Issues		
Use	For testing and explanation.	Use when sufficient theory exists to deduce generalizations between measured variables and latent constructs and infer the network among the latent constructs (including over-identifying assumptions). The results allow one to explain the phenomena including an explanation of strengths and weaknesses of the theory as a result of the test. Not ideal for exploratory research.
How to conduct a CFA	To get the full test of construct validity.	Avoid testing piecemeal and do not test using composites. Only when all constructs and variables are included can one adequately assess fit validity, convergent validity and discriminant validity.
Sample size	SEM not as sensitive to low sample size as once thought.	N = 100 is sufficient for most applications as long as measurement is good (AVE of .5 or better). Bottom line is that the generalizability necessary in the study is a more determinant criteria for sample size than statistical approach. No technique can make up for a sample that cannot generalize.
Estimation technique	Multiple linear options available.	Maximum likelihood estimation proves fairly robust and is applicable in most marketing research studies barring gross problems in measurement

Missing data treatment	For individual observations.	As long as missing data is minimal (below 5 percent), no practical difference in results will occur. Pairwise and listwise (casewise) deletions are options as EM imputation of missing data. If more than ten percent data are missing, EM imputation is necessary.
Measurement scale level	Type of scale measurement.	Generally, metric (at least interval) measures are presumed. However, dichotomous variables can be included when structured as dummy variables or with an augmented moment matrix.
Number of indicators	How many scale items are needed?	3 items insure that a construct is statistically identified. 4 items insure that the construct is over-identified. Thus, best to use a minimum of 3 items. If a measure consists of a single item and must be included in theoretical model, the loading and error variance term should be fixed rather than free. Likewise, while we recommend avoiding two-item scales, if they must be used, fix the loading estimates to identify.
Software	What to use?	All widely available software can produce reliable results. It's best to report what software was used in the analysis. More programs are available all the time including freeware. Users should use software that they are comfortable with.
Measurement model description	Measurement model is measurement theory.	Describe all indicators completely and be clear on what indicators were dropped to obtain final CFA results.
Individual item error-variance terms	How to treat?	A congeneric measurement model is a psychometrically sound representation. That means no correlated error variance terms. Generally, if correlated error terms are necessary to achieve good fit, the measurement is flawed.

Latent variable measurement mode		Reflectively measurement models typical; formative measurement possible with MIMIC type approach or by including a sufficient number of outcomes.
Statistical identification	For recursive (one-way causality) models.	In general avoid two item and single item measures. The best case scenario is to over-identify constructs by including 4 items or more as indicators of each construct. That means the net df for that construct is greater than 0. If fewer than three items are included with a construct, consider fixing the loading parameter(s). Nonrecursive models (2-way causal flow) are more complex to identify.
Theoretical identification	Lack of parsimony.	When every latent construct is connected to every other latent construct with a directional, causal path, the model is considered theoretically saturated. Every possible path is estimated. The result should fit the same as a standard CFA model. An assessment of theoretical is not possible.
Indicator loadings size		Standardized indicator loadings ≥ 0.70
Construct reliability		Construct/Composite reliability ≥ 0.70 (in exploratory research 0.60 to 0.70 is considered acceptable)
Convergent validity	How to establish evidence?	AVE ≥ 0.50 for each construct
Discriminant validity	How to establish evidence?	AVE of a latent construct = higher than the construct's highest squared correlation with any other latent construct
Fit validity / Overall fit	χ^2, df	ALWAYS REPORT final χ^2 value and final degrees of freedom! The results should be judged based on complexity of the model and sample size (see Hair et al. 2010). Simple models (with 5 or fewer df should exhibit insignificant p-

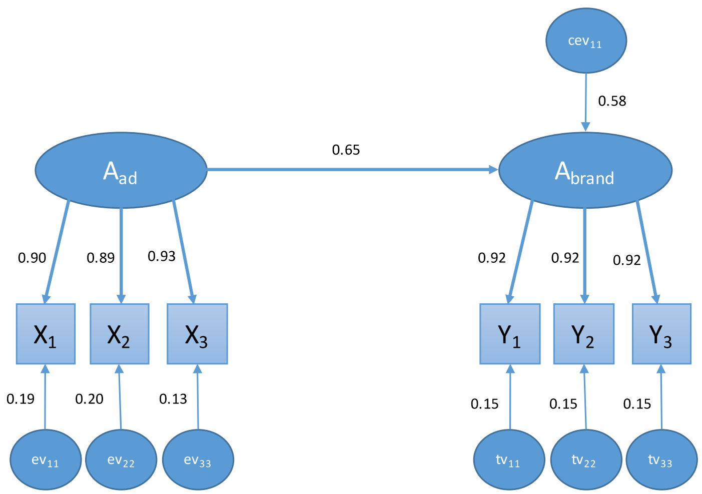
		values for good fit). The TFI, introduced above, provides a summary index of theoretical fit relative to measurement fit.
Overall validity	What else to report? Avoid a dump of all fit indices.	For CFA, in addition to χ^2 and df, report p-value, CFI and RMSEA to fully report fit. Also report AVE and CR (Construct/Composite Reliability).
Judging fit	How to know if fit is "good?"	No absolute standard exists for "good fit." Very simple models should be judged by the strictest criteria (insignificant p-values). More complex models should be judged by less strict criteria (CFI > .94; RMSEA < 0.08). See Hair et al. (2010) for tabled values. Comparing competing models with a χ^2 difference test is a valid way to give relative fit.
Formative Measurement Models		
Significance of weights		Report t-values and p-values
Multicollinearity		Examine for high correlations as indicators are theoretically independent.
Model Evaluation and Diagnostics		
Fit validity / Overall fit	How to know if theory fits?	Guidelines for CFA fit above apply. Researcher can use TFI to provide assessment of fit for the second step of the two-step SEM approach.
Path coefficient estimates		Assess significance and confidence intervals.
Number of indicators eliminated	When is analysis no longer confirmatory?	No more than 20% otherwise the CFA has become exploratory. That means if one starts with 20 items, no more than 4 can be dropped without admitting that the original measurement theory was flawed. In mid stages of scale

		development, developmental CFA may violate this rule as long as a subsequent validation adheres to it.
Mediation	How to model indirect effects?	Simply model the indirect relationships that the theory dictates. If model fits, then the evidence supports the indirect effects as modeled. Bootstrapping is available as an alternative to traditional t-values for parameter significance. Nonspecification bias can be diagnosed using error-variance terms.
Moderation	How to test moderation?	Although techniques exist to include multiplicative interactions, multiple group SEM is ideal for group level moderators based on the intuitive nature of the results and the ease of presentation of one group's effects relative to another group's effects.
Measurement invariance	Configural invariance	Means that the same factor structure can represent the theoretical latent constructs in each group. The evidence comes from a good fit for the multigroup CFA model.
Measurement invariance latent constructs	Metric invariance means that loadings do not vary by group and allow relationships to be compared between groups	Metric Invariance tests are necessary when the groups that are compared are sufficiently unique to suggest that they may not use the measures in a like manner. Such is typically the case when the groups have responded in different languages or are from different cultures. If the two groups are all from the same company, for instance, and all speak the same language, then that a-priori rationale for metric invariance may not be present. If in doubt, apply the test. Evidence comes when holding loadings constant between groups does not significantly diminish fit.

	Scaler invariance means that the intercept terms for each variable and construct do not vary by group and allow means to be compared between groups	Same as above except evidence comes when constraining intercept terms to be equal across groups does not diminish fit.
Indicators of problems	Parameter instability, implausibility, lack of convergence, incongruent df	Pay attention to these as they indicate problems with the data or errors in constructing the model.

FIGURE 1

Covariance, Structure, and Fit



		X1	X2	X3	Y1	Y2	Y3
	X1	2.6					
	X2	2.3	2.8				
S	X3	2.3	2.4	2.8			
	Y1	1.3	1.2	1.2	2		
	Y2	1.3	1.2	1.2	1.6	1.9	
	Y3	1.3	1.3	1.3	1.7	1.6	2
		X1	X2	X3	Y1	Y2	Y3
	X1	2.6					
	X2	2.3	2.8				
$\hat{\Sigma}$	X3	2.3	2.4	2.8			
	Y1	1.2	1.3	1.3	2		
	Y2	1.2	1.2	1.2	1.6	1.9	
	Y3	1.3	1.3	1.3	1.7	1.6	2
		X1	X2	X3	Y1	Y2	Y3
	X1	0					
	X2	0	0				
RES	X3	0	0	0			
	Y1	-0.1	0.1	0.1	0		
	Y2	-0.1	0	0	0	0	
	Y3	0	0	0	0	0	0

FIGURE 2

Frequency of *JA* Articles with SEM Terms by Year

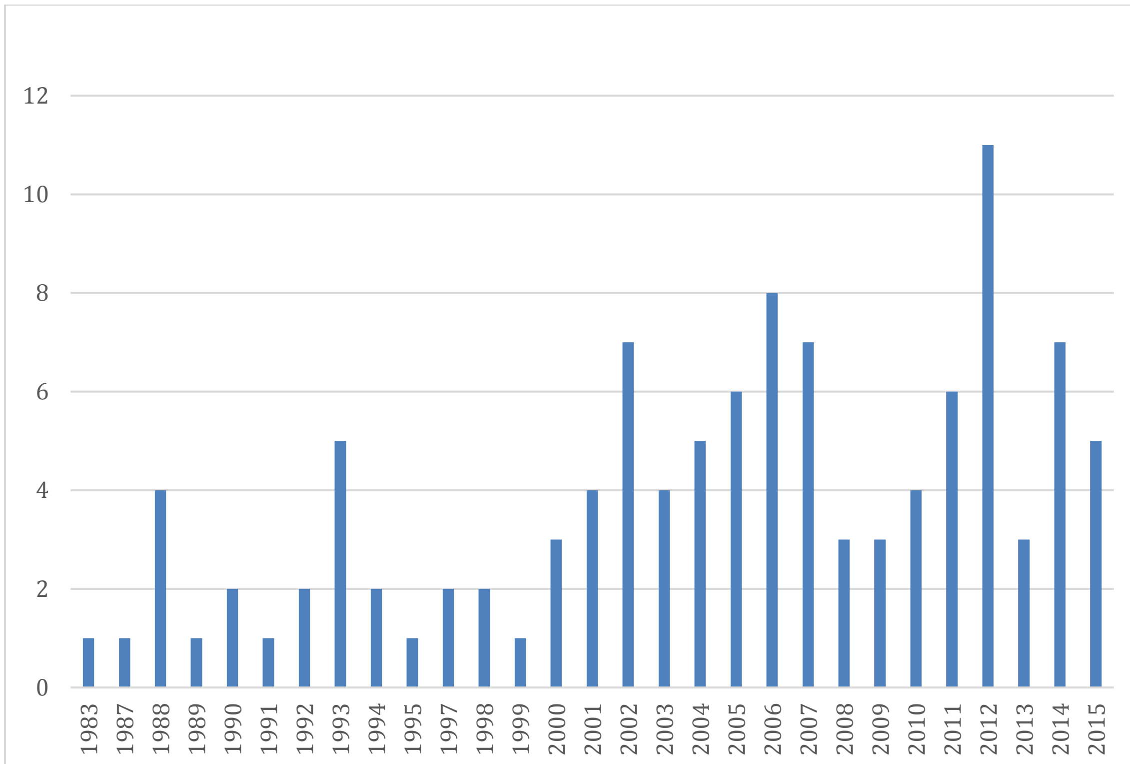


FIGURE 3

Software Usage in *JA* SEM Articles