

# Best Practice Recommendations for Using Structural Equation Modelling in Psychological Research

Todd G. Morrison\*, Melanie A. Morrison, Jessica M. McCutcheon

Department of Psychology, University of Saskatchewan, Saskatoon, Canada  
Email: \*Todd.Morrison@usask.ca

**How to cite this paper:** Morrison, T. G., Morrison, M. A., & McCutcheon, J. M. (2017). Best Practice Recommendations for Using Structural Equation Modelling in Psychological Research. *Psychology*, 8, 1326-1341.  
<https://doi.org/10.4236/psych.2017.89086>

**Received:** May 26, 2017  
**Accepted:** July 16, 2017  
**Published:** July 19, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Although structural equation modelling (SEM) is a popular analytic technique in the social sciences, it remains subject to misuse. The purposes of this paper are to assist psychologists interested in using SEM by: 1) providing a brief overview of this method; and 2) describing best practice recommendations for testing models and reporting findings. We also outline several resources that psychologists with limited familiarity about SEM may find helpful.

## Keywords

Structural Equation Modelling, Guidelines, Statistics, Primer

---

## 1. Introduction

By now, many psychologists will have encountered peer-reviewed papers in psychology and other disciplines that feature structural equation modelling (SEM). However, if you have not yet come across a paper that uses SEM or have heard reference to this statistical technique only in passing, you may be left with the following questions: What is SEM? Why does one use SEM? and What are SEM's key definitions and concepts? In our paper, we address these questions. We begin by providing a concise conceptual overview of SEM: its purpose, utility, and essential features, the latter through a diagrammatic representation of a mediational model. Identified next are criteria researchers should satisfy when using SEM as a statistical technique. We then close by highlighting various supplemental resources that may prove helpful to both novice and seasoned practitioners of SEM. Though there are certainly primers available on SEM, and excellent ones at that, we generated this contemporary, accessible, and interdisciplinary overview as a means of consolidating the most up-to-date recommenda-

tions possible in one place and, consequently, look to fuel new understanding and ongoing appropriate use of SEM as a data analytic method for the psychological community.

### 1.1. What Is Structural Equation Modelling?

SEM has been around for the past 60 years, but has increased significantly in popularity over the course of the last three decades (Von der Embse, 2016). SEM is a multivariate statistical technique that can be conceptualized as an extension of regression and, more aptly, a hybrid of factor analysis and path analysis (Weston & Gore, 2006). Though it is a complex method of data analysis, the beauty of SEM is that it allows a researcher to analyse the interrelationships among variables (akin to a factor analytic approach) and test hypothesized relationships among constructs (akin to a path analytic approach). Von der Embse (2016) further emphasizes that SEM enables testing of hypothesized relationships that are not possible with traditional data analytic methods. For instance, when using regression analyses, one must take a “step-by-step” approach to test interrelationships. With SEM, users are permitted to test a number of interrelationships simultaneously.

Since SEM often assumes linear relationships, it is similar to common statistical techniques such as analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), and multiple regression; yet, where SEM departs from the aforementioned is in its capacity to estimate and test complex patterns of relationships at the construct level. Weston & Gore (2006: p. 723) emphasize that, “unlike other general linear models, where constructs are represented by only one measure and measurement error is not modeled, SEM allows a researcher to use multiple measures to represent constructs and addresses the issue of measure-specific error.” According to these authors, it is this difference that allows one to test the construct validity of factors. With respect to measurement-specific error (i.e., error produced via multiple raters, administrations, or test variations), the measurement error that typically accompanies each observed variable is taken into account and appears in the form of measurement error variables. Thus, conclusions researchers may draw about relationships between constructs when using SEM are not biased by measurement error, as these relationships “are equivalent to relationships between variables of perfect reliability” (Werner & Schermelleh-Engel, 2009, p. 1). In all, SEM departs from other statistical methods because it enables researchers to include multiple measures and reduce their measurement error—error inherent in any data utilized in the social sciences or related disciplines.

When testing the interrelationships among variables and constructs, as one does with SEM, researchers should be aware that they are, in essence, taking a confirmatory (i.e., hypothesis-testing) approach rather than an exploratory approach to their data analysis (Byrne, 2016). A confirmatory approach is adopted because researchers specify *a priori* the interrelationships that are theorized to exist (i.e., through specification of a model), with the next step being to test how

well the theorized model fits the obtained (sample) data. Confirmation of fit in this instance can be assessed at a global level (i.e., the theoretical model does or does not fit the data), local level (i.e., the model reproduces or does not reproduce hypothesized relationships between specific variables), and an exploratory level (i.e., determining which aspects of the model require improvement; Von der Embse, 2016).

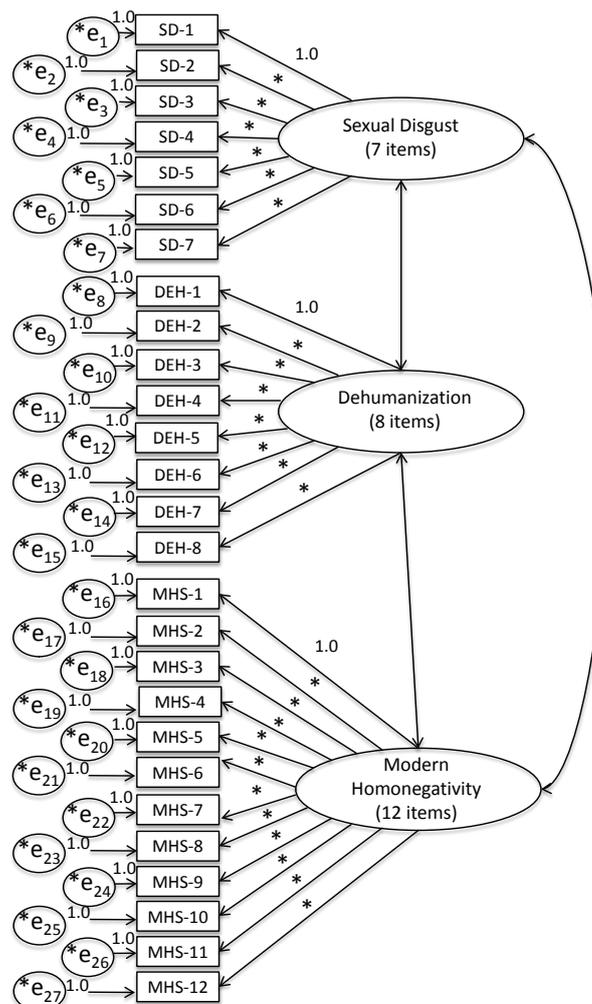
## 1.2. What are Essential Features of Structural Equation Modelling?

A structural equation (SE) model is a complex amalgam of hypothesized interrelationships represented and tested with equations (Von der Embse, 2016). Specifically, SEM includes, describes, and tests the interrelationships between two types of variables: manifest and latent. *Manifest* variables are those that can be directly observed (measured); alternatively, *latent* variables are those that cannot be observed (measured) directly due to their abstract nature (Byrne, 2016; Xiong, Skitmore, & Xia, 2015). Latent variables can be further broken down into *exogenous* and *endogenous* variables. According to Byrne (2016), exogenous latent variables are analogous to independent variables in that they are thought to “cause” fluctuations in the values of other latent variables in the model. Importantly, exogenous variables are not explained by the SE model, but rather can be viewed as factors external to the model that create changes in the value of exogenous variables (Byrne, 2016). Variables such as gender and socioeconomic status may be factors external to a SE model and may produce fluctuations in other latent variables. Endogenous variables, on the other hand, are synonymous with dependent variables. Endogenous latent variables are influenced by exogenous latent variables and this influence can occur directly or indirectly. Since all latent variables influencing endogenous variables are specified within a SE model, any change in the value of endogenous latent variables is thought to be explained by the model (Byrne, 2016; Winke, 2014).

A SE model typically consists of a *measurement model*, which is a set of observed variables that represent a small number of latent (unobserved) variables. The measurement model describes the relationship between observed variables (e.g., instruments) and unobserved variables; that is, it connects the instruments that are used to the constructs they are hypothesized to measure (Byrne, 2016; Weston & Gore, 2006). Confirmatory factor analysis (CFA) would then be employed as a means of determining the pattern of loadings of each newly emerging hypothesized factor. A SE model also typically consists of a *structural model*, which is a schematic depicting the interrelationships among latent variables (Von der Embse, 2016). When the measurement and structural models are considered together, the model is called the full or complete structural model (Weston & Gore, 2006). The complete structural model allows researchers to specify regression structures among the latent variables, wherein “a structural model that specifies direction of cause from one direction only is called a *recursive model*, and one that allows for reciprocal or feedback effects is termed a *nonre-*

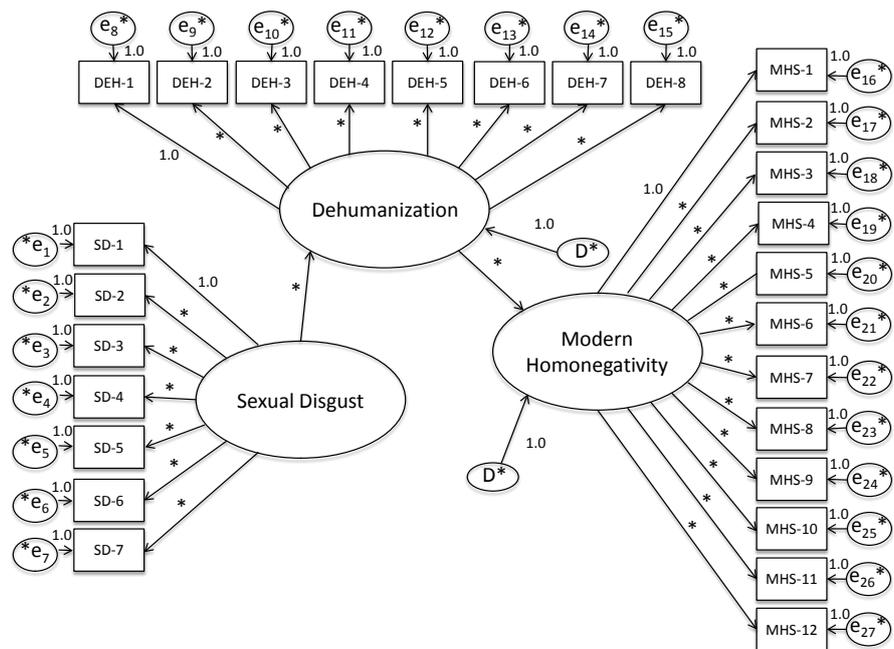
*cursive model*” (Byrne, 2016, p. 7).

**Figure 1** depicts an example of a measurement model. In this model, each of the three latent variables—sexual disgust, dehumanization, and modern homonegativity—are symbolized by ellipses (ovals). The observed or manifest variables, referred to as *indicators* here, appear in rectangles. Sexual disgust has seven observed (measured) variables (SD1-SD7; Tybur, Lieberman, & Griskevicius, 2009), dehumanization has eight observed (measured) variables (DEH1-DEH8; Bastian, Denson, & Haslam, 2013), and modern homonegativity has 12 observed (measured) variables (MHS1-MHS12; Morrison & Morrison, 2003). For practical purposes, only one measure hypothesized to represent each latent variable is shown; but, ideally, there would be several measures indicated for each latent variable (Weston & Gore, 2006). Of note, too, is that each manifest variable (each individually observed item in this case) has its own error term (Byrne, 2016). As Weston and Gore (2006) state, measures that are reliable and have fewer errors will be “better indicators of their respective latent variable, much as the items in a scale that most accurately represent the underlying construct have the highest factor loadings in a factor analysis” (p. 726).



**Figure 1.** Confirmatory factor analysis used to test measurement model.

**Figure 2** provides a schematic of a fully mediated complete model. In this model, it is hypothesized that dehumanization mediates the effects of sexual disgust on modern homonegativity (i.e., negative beliefs about gay men). Specifically, greater levels of sexual disgust are associated with greater levels of dehumanization which, in turn, is associated with greater levels of homonegativity toward gay men. The interrelationships amongst latent variables can be conceptualized as covariances, direct effects, or indirect (mediated) effects. Covariances are similar to correlations, and, if present, would be symbolized in **Figure 2** with double-headed arrows. Double-headed arrows in any SE model signify the presence of non-directional relationships (covariances) between latent variables. As there were no double-headed arrows appearing in **Figure 2**, non-directional relationships were not expected. **Figure 2** does, however, indicate the presence of hypothesized direct relationships between sexual disgust and dehumanization, and dehumanization and modern homonegativity. These direct effects are symbolized by single-headed arrows. As **Weston and Gore (2006)** point out, though directional claims are being made regarding the interrelationships amongst latent variables, the relationships themselves are not causal. Interpreting the strength of the relationships should be carried out much akin to the way one would interpret regression weights. Also in **Figure 2**, each latent construct is represented by a series of observed indicator variables, each with its own error term (i.e., \*e, in the present schematic). The error term for the latent variables is referred to as disturbance and is symbolized with a *D* (*D\** in the Figure). Finally, regarding the delineation of exogenous and endogenous variables, of the three latent variables pictured, only sexual disgust is independent (i.e., not predicted by any other latent variable). Dehumanization and modern homonegativity are



**Figure 2.** Fully mediated complete model. *Note:* Asterisks represent parameters to be estimated.

endogenous because they are dependent on (i.e., predicted by) their respective latent variables. Given that key concepts related to SEM have been described, we now provide an overview of the practices that researchers are encouraged to employ when using SEM.

## 2. Best Practices before Testing a Model

### 2.1. Model Development

When formulating a model, a critical issue pertains to the number of manifest indicators that one should have for each latent variable. The consensus is that  $\geq 2$  indicators per latent variable is required. In terms of an upper limit, however, no consistent recommendation emerges. Ping (2008) notes an apparent ceiling of six indicators per latent variable due to “extensive item weeding” that may be attributable, in part, to “persistent model-to-data fit difficulties” (p. 2). Our recommendation is that one take into consideration the sometimes competing interests of parsimony and model thoroughness. Finally, with respect to manifest indicators, Ho’s (2013) advice is sound: “researchers should be guided by the axiom that it is preferable to employ a relatively small number of good indicators than to delude oneself with a relatively large number of poor ones” (p. 432). It is essential that measures selected to represent latent variables be psychometrically robust. Of particular importance are the matters of content validity, scale score reliability, and construct validity.

Content validity may be defined as the relevance and representativeness of the targeted construct, across all features of a measure (e.g., the scale items and the instructions provided to respondents: Haynes, Richard, & Kubany, 1995). This type of validity may be established in the following ways: a) conducting an extensive review of the literature pertinent to the construct (including published and unpublished work); b) consulting with stakeholders from relevant groups that are able to furnish valuable insights about the construct; and c) using experts to gauge the suitability of all items designed to measure the construct (Yaghmaie, 2009). If there are insufficient details about an instrument’s content validity, then we recommend researchers opt for another measure.

In terms of scale score reliability, the most popular estimate is Cronbach’s alpha, which is the “expected correlation between an actual test and a hypothetical alternative form of the same length” (Carmines & Zeller, 1979: p. 45). As reliability is a product of scale scores, it must be calculated whenever a researcher intends to average or sum a multi-item measure (Streiner, 2003). A Cronbach’s alpha coefficient of .80 often serves as the cut-off for “good” reliability, with Streiner (2003) advocating a maximum value of .90. (Values exceeding .90 suggest item redundancy.) However, we do not advise rigid adherence to cut-off values, as there may be instances where low alpha coefficients are defensible (see Johns & Holden, 1997; Schmitt, 1996).

It should be noted that Cronbach’s alpha has been subject to considerable criticism and that other forms of scale score reliability have been recommended such as Omega (e.g., Dunn, Baguley, & Brunnsden, 2014). For example, Peters

(2014) notes that Cronbach's alpha uses the essentially tau-equivalent model which operates in accordance with a specific set of assumptions; ones that are seldom met with real-world psychological data. These assumptions include: 1) all items measure the same underlying variable; 2) all items are of comparable strength in terms of their association with that underlying variable; 3) unidimensionality; and 3) item variances and covariances are equal (Peters, 2014). For these reasons, we, subsequently, describe other indicators of reliability that practitioners of SEM may wish to test.

Carmines and Zeller (1979) note that there are two principal forms of construct validity: convergent and discriminant. Convergent validity examines whether scores on the measure that is being validated correlate with other variables with which, for theoretical and/or empirical reasons, they should be correlated. Discriminant validity, on the other hand, targets variables that, again for theoretical and/or empirical reasons, should have a negligible association with the measure being validated (Springer, Abell, & Hudson, 2002).

Testing a measure's psychometric soundness is an iterative process that necessitates the accumulation of multiple strands of validation across diverse samples. We recommend that, when targeting measures for inclusion in a model that will be tested with SEM, researchers review source articles that detail the precise steps used to create and refine a scale's items as well as the tests conducted to evaluate scale score reliability and validity. Finally, it is vital to emphasize that utilizing instruments that seem to be psychometrically robust does not preclude assessing the reliability and validity of the measurement models (i.e., the confirmatory factor components of a SE model). We review these topics later in the document.

## 2.2. Alternative Models

It is important to acknowledge *a priori* the existence of models that are rivals to the one being tested (Weston & Gore, 2006). Such rivals may reflect "other theoretical propositions and/or contradictions in empirical findings" (Nunkoo, Ramkissoon, & Gursoy, 2013: p. 761) and should be made explicit and tested.

## 2.3. Sample Size Requirements

The issue of how many participants are needed to use SEM as an analytic technique remains a point of contention (see, for example, Barrett, 2007; Iacobucci, 2010). However, rules-of-thumb do not appear to be appropriate as issues such as model complexity, amount of missing data, and size of factor loadings have implications for the numbers of participants required (Wolf, Harrington, Clark, & Miller, 2013). Using Monte Carlo simulation, Wolf et al. found that if a researcher wanted to conduct a confirmatory factor analysis with a single latent variable and 6 indicators (having average loadings of .65), a sample size of 60 was adequate. However, for a more complex mediation model, consisting of three latent factors (each having three manifest indicators), the minimum sample needed was 180. To assist with sample size decision-making, an *a priori* sam-

ple size calculator may be helpful (see, for example: <http://www.danielsoper.com/statcalc/calculator.aspx?id=89>). With this calculator, individuals must provide the anticipated effect size (typically .1 to .3), the desired level of statistical power (usually set at .80), the number of latent variables included in the model, the number of manifest indicators (i.e., measured variables), and the probability value used to denote “statistical significance” (traditionally .05).

### 3. Best Practices When Testing Models

#### 3.1. Data-Related Assumptions

The most commonly used estimation method in SEM is maximum-likelihood (ML). ML has various assumptions including: 1) there will be no missing data; and 2) the endogenous (or dependent) variables will have a multivariate normal distribution.

If a complete datafile is unavailable, then the researcher must test whether the data are missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). If data are MCAR, then the “missingness” on the variable of interest is unrelated to any of the variables in the dataset. If the data are MAR, then systematic differences may exist between missing and observed values; however, these differences are accounted for by *other* variables in the dataset (Bhaskaran & Smeeth, 2014). Finally, if data are MNAR, then there is a systematic pattern to the missing data (the presence or absence of a score on variable X is related to the variable itself). To determine whether data are MCAR, Little’s MCAR test can be used (i.e., a statistically non-significant  $p$  value [ $>.05$ ] denotes that data are MCAR). If Little’s test is statistically significant, then data may be MAR or MNAR; further investigation of participants with missing data is required. If data are MCAR or MAR, the sample is large, and the proportion of missing data is modest ( $< 5\%$ ), listwise deletion is a reasonable option (Green, 2016). An alternative approach is using Multiple Imputation (MI) to estimate missing data. Finally, when data are MNAR, item parcels may be useful (see Orcan, 2013); though for the novice practitioner, SEM would not be recommended (Allison, 2003).

Having data that are multivariate normal is a key assumption when performing SEM (using the ML default). Although univariate normality does not guarantee the multivariate normality of one’s data, we recommend that each variable be scrutinized to identify any deviations from a normal distribution. Ghasemi and Zahediasl (2012) provide a straightforward overview of the primary visual and statistical techniques that may be used to gauge univariate normality. The two key considerations are skew and kurtosis. Skewness refers to the *lack* of symmetry in the distribution of one’s data (i.e., for a symmetrical distribution, or one without skew, the distribution to the left or right of the center-point looks identical: Field, 2013). Kurtosis may be thought of as the “tail-heaviness” of the distribution of one’s data (i.e., leptokurtosis happens when the number and extremity of outliers is smaller than would occur with a normal distribution; pla-

tykurtosis occurs when the number and extremity of outliers is greater than would take place with a normal distribution). Suggested cut-offs for the skewness index (i.e., skew divided by the standard error of skew) and the kurtosis index (i.e., kurtosis divided by the standard error of kurtosis) are absolute values greater than 3 and 10, respectively (Weston & Gore, 2006). Determining multivariate normality is more difficult, as popular statistical packages such as SPSS do not offer formal tests of multivariate skewness and kurtosis. However, Wan Nor (2015) offers a step-by-step guide to graphically assessing multivariate normality using SPSS and DeCarlo provides SPSS syntax that may be used to determine both univariate and multivariate normality (see: <http://www.columbia.edu/~ld208/>). If data are non-normal, they may be transformed. Tabachnick and Fidell (2006) provide SPSS and SAS compute commands to address issues of moderate to severe positive and negative skew (see page 89). Another option is to assess model fit using a  $p$  value that is not ML-based (e.g., Bollen-Stine in AMOS).

### 3.2. Two-Stage Modelling

When conducting SEM, it is recommended that the measurement models be assessed first, using confirmatory factor analysis (CFA), followed by simultaneous assessment of the measurement and structural models (Anderson & Gerbing, 1988). As noted earlier, each measurement model consists of at least one latent factor, its measured indicators and their associated error terms. The structural model represents the predicted associations among the latent variables based on theory and/or prior empirical research (Xiang et al., 2015). Thus, a model containing two latent variables (Y1 and Y2), each of which is represented by three manifest indicators (Y1: x1, x2, x3; Y2: x1, x2, x3) would consist of two measurement models (one for Y1 and one for Y2) and one structural model that tests Y1 and Y2 simultaneously. With the two-stage approach, each measurement model is tested. If adequate fit is not obtained, then each model may be subject to re-specification, provided one can justify doing so on the basis of theory, indicator content, and/or past research (Anderson & Gerbing, 1988). It should be noted that, unless a compelling reason is specified *a priori*, simply correlating error terms to improve fit is not recommended because doing so takes “advantage of chance, at a cost of only a single degree of freedom, with a consequent loss of interpretability and theoretical meaningfulness” (Anderson & Gerbing, 1988: p. 417). The structural model then is evaluated.

### 3.3. Reliability and Validity

When testing each measurement model, using confirmatory factor analysis, output can be used to assess indicator and composite reliabilities as well as convergent and discriminant validities. Indicator reliability (IR) refers to the proportion of variance in each measured variable that is accounted for by the latent factor it supposedly represents (O'Rourke & Hatcher, 2013). Calculating IR is straightforward as it merely involves squaring the standardized factor loading

for each measured variable (O'Rourke & Hatcher, 2013). Thus, if latent variable Y had three indicators (x1, x2, and x3) with factor loadings of .54, .67, and .80, respectively, IR coefficients would be .29 (.54<sup>2</sup>), .45 (.67<sup>2</sup>), and .64 (.80<sup>2</sup>). Note that the IR values for x1 and x2 are low and may warrant scrutiny. Composite reliability (CR), which may be viewed as analogous to Cronbach's alpha coefficient, also should be computed for each latent factor. The following steps may be used to compute CR: a) calculate IR for each item (i.e., each factor loading squared); b) determine the error variance for each item by subtracting each IR value from 1 (i.e., 1-IR); c) for a given latent variable, sum the standardized factor loadings and then square the sum; and d) for a given latent variable, take the squared sum of the factor loadings ( $\Sigma\text{SSL}$ ) and divide that number by itself ( $\Sigma\text{SSL}$ ) plus the sum of the error variance ( $\Sigma\text{EV}$ ); that is:  $\Sigma\text{SSL}/\Sigma\text{SSL} + \Sigma\text{EV}$ . The resultant value denotes the CR for the latent variable in question. Using the hypothetical values listed above (i.e., IRs for x1, x2, and x3 = .29, .45, and .64, respectively), the error variances are:  $1 - .29 = .71$  for x1;  $1 - .45 = .55$  for x2; and  $1 - .64 = .36$  for x3. As noted earlier, the factor loadings were .54, .67, and .80. The sum of these values squared is 4.04 (i.e.,  $.54 + .67 + .80 = 2.01^2$ ). The sum of the error variances is 1.62 (i.e.,  $.71 + .55 + .36$ ). Thus, the resultant CR for latent variable Y is .71 (i.e.,  $4.04/4.04 + 1.62$ ). As values of .70+ are considered to be acceptable in research that is not strictly exploratory (Nunkoo et al., 2015), this hypothetical CR value is satisfactory.

The average variance extracted (AVE) may be used to test the convergent validity of the measurement model. To compute AVE for a given latent variable, simply square each standardized factor loading, sum them, and divide by the total number of loadings. Using the aforementioned hypothetical loadings (.54, .67, and .80), the squared sum is 1.38 ( $.54^2 + .67^2 + .80^2$ ); dividing that total by 3 (number of loadings), the AVE is .46. This value is below the typical cut-off used to establish convergent validity (.50+; Nunkoo et al., 2015). Provided that one has no more than 10 measured indicators per latent factor, the following online calculator may be useful when wishing to determine AVE:

<http://www.watoowatoo.net/sem/sem.html>. This calculator also provides composite reliability coefficients (see: Jöreskog's rho).

Finally, to assess discriminant validity, the procedure outlined by Fornell & Larcker (1981) appears to be reasonable. Using latent variables  $Y_1$  and  $Y_2$  as hypothetical examples, the researcher would first calculate AVE values for the two variables and then contrast these values with the squared correlation between  $Y_1$  and  $Y_2$ . If *both* AVE numbers are greater than the square of the correlation, discriminant validity has been demonstrated.

### 3.4. Model Fit

A broad range of fit indices, encompassing four broad categories (i.e., overall model fit, incremental fit, absolute fit, and predictive fit), should be used (Worthington & Whittaker, 2006). Overall model fit, which includes the chi-square

test, tests precisely what it describes: whether the model fits the observed data. Ropovik (2015) notes that, while a statistically significant chi-square value is often ignored on the grounds that the test itself is overly sensitive when large samples are used, the “only message that a significant  $\chi^2$  tells is... take a good look at that model [as] something may be wrong here” (p. 4). Further, the attainment of fit using other indices (e.g., GFI or RMSEA) does not necessarily mean that the chi-square test was statistically significant because of a trivial misspecification. Detailed analysis of the model is required.

Incremental fit indices compare the model that is being tested to a baseline model which, typically, is one in which all variables are uncorrelated (Worthington & Whittaker, 2006). Sample indices include: the normed fit index (NFI), the comparative fit index (CFI), and the Tucker Lewis index (TLI). Absolute fit indices, such as the root mean square error of approximation (RMSEA), goodness-of-fit index (GFI), and the standardized root mean square residual (SRMR), determine how well a model specified *a priori* reproduces the sample data (Hooper, Coughlan, & Mullen, 2008). If the SRMR is not reported, then we recommend researchers furnish a table of correlation residuals, which represent the difference between a correlation for the model and an observed correlation. The greater the absolute magnitude of a given correlation residual, the greater the misfit between the model and the actual data for the two variables in question.

With respect to cut-off values for various fit indices, the current perspective is that individuals should avoid mindlessly using cut-off values and that “no single cut-off value for any particular [fit index] can be broadly applied across latent variable models” (McNeish, An, & Hancock, 2017: p. 8). Measurement quality, which McNeish et al. operationalize as the magnitude of the standardized loadings between each latent construct and its manifest variables, plays a critical role with respect to the interpretability of cut-off values. Referring to the reliability paradox, these researchers note that fit indices tend to be worse when measurement quality is higher rather than lower. Thus, a model with standardized loadings of .90 may produce worse fit statistics than a model with standardized loadings of .40—although the former has better data-model fit than does the latter.

Finally, predictive fit indicators examine “how well the structural equation model would fit other samples from the same population” (Worthington & Whittaker, 2006: p. 828). One common example is the Akaike Information Criterion (AIC), which measures “badness” of fit (i.e., the model with the *lowest* AIC value is the most parsimonious and, thus, would be chosen: Schermelleh-Engel, Moosbrugger, & Müller, 2003).

#### 4. Reporting Guidelines for SEM

When writing a manuscript that involves SEM, various pieces of information are essential if readers are to make an informed decision about the appropriateness of the findings. We recommend the following be reported:

1. As determined by an *a priori* power analysis, the minimum number of participants needed, given the models that are being tested.

2. At least one alternative model that is plausible in light of extant theory or relevant empirical findings.
3. Graphical displays of all measurement and structural models.
4. Brief details about the psychometric properties of scale scores for all measured variables (e.g., Cronbach's alpha and its 95% confidence intervals or, preferably, omega as well as 2 to 3 sentences per measure detailing evidence of content and construct validities).
5. The proportion of data that are missing and whether missing data are MCAR, MAR, or MNAR. As well, researchers should explicate how this decision was reached (e.g., why does a researcher assume missing data are MAR?), and the action taken to address missing data.
6. Assessments of univariate and multivariate normality for all measured indicators.
7. The estimation method used to generate all SEMs (default is ML estimation).
8. The software (including version) that was used to analyze the data.
9. In accordance with the advised two-step approach, full CFA details about each measurement model followed by complete SEM details about the structural model.
10. Indicator and composite reliabilities.
11. Average variance extracted (AVE) for each latent factor which denote convergent validity.
12. Discriminant validity of latent factors, as per [Fornell and Larcker's \(1981\)](#) test.
13. All standardized loadings from latent variables to manifest variables (reflective models).
14. Fit indices that reflect overall, absolute, and incremental fit. If applicable, predictive fit indicators should be included.
15. A clear and compelling rationale for all post-hoc model modifications.
16. An indicator of effect size for the final model.

## 5. Useful Resources

We would like to conclude this brief primer by listing resources that we recommend both novice and experienced practitioners of SEM consult.

1. [Byrne, B. M. \(2016\)](#). *Structural equation modelling with AMOS: Basic concepts, applications, and programming* (3<sup>rd</sup> ed.). New York: Routledge.

The popularity of AMOS software for SEM analysis makes Byrne's (2016) book a valuable resource for many SEM users. Byrne provides an easy-to-understand introduction to SEM and AMOS, not requiring the reader to have any pre-existing knowledge about SEM or any software programs. She includes detailed instructions on calculating reliability and validity (a best practice that has largely been ignored by researchers), drop-down menus, charts, and tables directly from AMOS, which allows the reader to follow along without any difficulty. Moreover, the data that are used in the examples are available to the readers online, allowing them to fully ensure they can conduct SEM using AMOS before

they try with their own data.

2. Gaskin, J. [James Gaskin]. (2014, May 8). *SEM BootCamp 2014 Series* [Video Files]. Retrieved from [https://www.youtube.com/watch?v=C\\_Jf4l0PF18](https://www.youtube.com/watch?v=C_Jf4l0PF18)

Dr. James Gaskin, from Brigham Young University, offers a YouTube series, titled “SEM BootCamp,” that takes the viewers through best practices for conducting SEM using AMOS. Topics include, but are not limited to, data screening, assumption testing, mediation and moderation, and potential issues that might be encountered. The videos provide a user-friendly and step-by-step guide, emphasizing both theory and practice that would be very helpful to those who are novices in SEM. Additionally, viewers are able to access the data files he uses in his examples, allowing them to follow along through the guided examples.

3. Researchgate.net

This website facilitates communication from academics across the globe and, thus, provides an invaluable source of information about all facets of SEM. All one needs to do is “Google” a specific question and, in conjunction with the word “researchgate,” a discussion containing relevant information and resources will emerge. For example, using the search terms “multivariate normality,” “SEM,” and “researchgate” produced 4,650 results (as of February 26, 2017). These hits included discussions about what steps should be taken to test for multivariate normality; what can be done if this assumption is violated; and whether specific software were better suited to address non-normality.

4. O'Rourke & Hatcher (2013). *A step-by-step approach to using SAS for factor analysis and structural equation modelling*. SAS Institute.

Even for non-SAS practitioners, this book offers an accessible overview of SEM by using straightforward language and clear examples. For instance, the authors provide an illustrated, step-by-step guide for computing indicator and composite reliabilities as well as convergent and discriminant validities for latent factors.

5. Winke (2014). Testing hypotheses about language learning using structural equation modelling. *Annual Review of Applied Linguistics*, 34, 102-122.

Dr. Paula Winke has written a paper that provides an excellent introduction to SEM for the novice user. Winke provides examples from applied linguistics that a researcher can understand, and she has a keen ability to describe SEM in a very clear manner. Her overview of what is contained in SEM models (both measurement and structural) are accessible and manage to inspire the reader rather than discourage.

## 6. Conclusion

SEM is a powerful statistical technique; one that permits assessing “latent variables at the observation level (i.e., a measurement model) and testing hypothesized relationships between latent variables at the theoretical level (i.e., a structural model)” (Nunkoo et al., 2013: p. 759). However, like any statistical procedure, SEM can be subject to inappropriate and indiscriminant use. To maximize

its value in psychological research, it is essential that psychologists should be informed practitioners of SEM. By outlining best practice recommendations that should be followed both prior to, and during, model testing as well as elucidating supplemental resources about SEM that we have found to be valuable, we hope this paper will encourage improved use of this analytic technique.

## Acknowledgements

This study was conducted with the support of a Social Sciences and Humanities Research Council of Canada (SSHRC) Insight Grant (#346011) awarded to the first and last authors.

## References

- Allison, P. D. (2003). Missing Data Techniques for Structural Equation Modeling. *Journal of Abnormal Psychology, 112*, 545-557. <https://doi.org/10.1037/0021-843X.112.4.545>
- Anderson, J. C., & Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin, 103*, 411-423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Barrett, P. (2007). Structural Equation Modeling: Adjudging Model Fit. *Personality and Individual Differences, 42*, 815-824. <https://doi.org/10.1016/j.paid.2006.09.018>
- Bastian, B., Denson, T. F., & Haslam, N. (2013). The Roles of Dehumanization and Moral Outrage in Retributive Justice. *PloS One, 8*, e61842. <https://doi.org/10.1371/journal.pone.0061842>
- Bhaskaran, K., & Smeeth, L. (2014). What Is the Difference between Missing Completely at Random and Missing at Random? *International Journal of Epidemiology, 43*, 1336-1339. <https://doi.org/10.1093/ije/dyu080>
- Byrne, B. M. (2016). *Structural Equation Modelling with AMOS: Basic Concepts, Applications, and Programming* (3rd ed.). New York: Routledge.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment* (Vol. 17). Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781412985642>
- Dunn, T. J., Baguley, T., & Brunson, V. (2014). From Alpha to Omega: A Practical Solution to the Pervasive Problem of Internal Consistency Estimation. *British Journal of Psychology, 105*, 399-412. <https://doi.org/10.1111/bjop.12046>
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Thousand Oaks, CA: Sage.
- Fornell, C., & Larcker, D.F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research, 18*, 39-50. <https://doi.org/10.2307/3151312>
- Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism, 10*, 486-489. <https://doi.org/10.5812/ijem.3505>
- Green, T. (2016). A Methodological Review of Structural Equation Modelling in Higher Education Research. *Studies in Higher Education, 41*, 2125-2155. <https://doi.org/10.1080/03075079.2015.1021670>
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment, 7*, 238-247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Ho, R. (2013). *Handbook of Univariate and Multivariate Data Analysis with IBM SPSS*.

- Boca Raton, FL: CRC Press. <https://doi.org/10.1201/b15605>
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6, 53-60. <http://www.ejbrm.com>
- Iacobucci, D. (2010). Structural Equation Modeling: Fit Indices, Sample Size and Advanced Topics. *Journal of Consumer Psychology*, 20, 90-98.
- McNeish, D., An, J., & Hancock, G. R. (2017). The Thorny Relation between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. *Journal of Personality Assessment*, 1-10.
- Morrison, M. A., & Morrison, T. G. (2003). Development and Validation of a Scale Measuring Modern Prejudice toward Gay Men and Lesbian Women. *Journal of Homosexuality*, 43, 15-37. [https://doi.org/10.1300/j082v43n02\\_02](https://doi.org/10.1300/j082v43n02_02)
- Nunkoo, R., Ramkissoon, H., & Gursoy, D. (2013). Use of Structural Equation Modeling in Tourism Research: Past, Present, and Future. *Journal of Travel Research*, 52, 759-771. <https://doi.org/10.1177/0047287513478503>
- O'Rourke, N., & Hatcher, L. (2013). *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling*. SAS Institute.
- Orcan, F. (2013). *Use of Item Parceling in Structural Equation Modeling with Missing Data*. Unpublished Doctoral Dissertation, Florida State University.
- Peters, G. J. Y. (2014). The Alpha and the Omega of Scale Reliability and Validity: Why and How to Abandon Cronbach's Alpha and the Route towards More Comprehensive Assessment of Scale Quality. *European Health Psychologist*, 16, 56-69.
- Ping, R. (2008). *On the Maximum of about Six Indicators per Latent Variable with Real-World Data*. Unpublished Document. <http://www.wright.edu/~robert.ping/puzzle6.doc>
- Ropovik, I. (2015). A Cautionary Note on Testing Latent Variable Models. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01715>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment*, 8, 350-353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Springer, D. W., Abell, N., & Hudson, W. W. (2002). Creating and Validating Rapid Assessment Instruments for Practice and Research: Part 1. *Research on Social Work Practice*, 12, 408-439. <https://doi.org/10.1177/1049731502012003005>
- Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, 80, 99-103. [https://doi.org/10.1207/S15327752JPA8001\\_18](https://doi.org/10.1207/S15327752JPA8001_18)
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using Multivariate Statistics* (5th ed.). Boston: Allyn & Bacon.
- Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, Mating, and Morality: Individual Differences in Three Functional Domains of Disgust. *Journal of Personality and Social Psychology*, 97, 103-122. <https://doi.org/10.1037/a0015474>
- Von der Embse, N. P. (2016). What School Psychologists Need to Know about Structural Equation Modelling. *School Psychologists as Consumers of Research*, 44, 10-12.
- Wan Nor, A. (2015). The Graphical Assessment of Multivariate Normality Using SPSS. *Education in Medicine Journal*, 7. <http://www.eduimed.com/index.php/eimj/article/view/361>

<https://doi.org/10.5959/eimj.v7i2.361>

Weston, R., & Gore, P. A. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist, 34*, 719-751.

<https://doi.org/10.1177/0011000006286345>

Winke, P. (2014). Testing Hypotheses about Language Learning Using Structural Equation Modelling. *Annual Review of Applied Linguistics, 34*, 102-122.

<https://doi.org/10.1017/S0267190514000075>

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement, 73*, 913-934.

<https://doi.org/10.1177/0013164413495237>

Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist, 34*, 806-838. <https://doi.org/10.1177/0011000006288127>

Xiong, B., Skitmore, M., & Xia, B. (2015). A Critical Review of Structural Equation Modelling Applications in Construction Research. *Automation in Construction, 49*, 59-70.

Yaghmaie, F. (2009). Content Validity and Its Estimation. *Journal of Medical Education, 3*, 25-27.



Scientific Research Publishing

**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [psych@scirp.org](mailto:psych@scirp.org)