PRESENTING SEM MODELS: JOURNAL AND CONFERENCE CONSIDERATIONS

John R Hamilton*, James Cook University, Cairns, QLD, Australia, 4870.
Ph +(61)740421091 John.Hamilton@jcu.edu.au
Singwhat Tee, James Cook University, Cairns, QLD, Australia, 4870.
Ph +(61)740421494 Singwhat.Tee@jcu.edu.au

ABSTRACT

SEM is a versatile statistical technique for research across disciplines. The materials presented in SEM studies offer differing degrees of detail. This creates difficulties in assessing the validity and the applications of such SEM studies. This paper provides a working checklist of SEM required quality indicators. This checklist and five additional fixes are designed to assist SEM researchers and to help the reviewing of SEM studies offered for publication.

Structural equation modelling, quality, model fit indicators

INTRODUCTION

Structural Equation Modelling (SEM) is today a key quantitative analysis research technique of choice for industry and for researchers across disciplines.
Establishing the best model is guided by the processes within the SEM build. Here the model should represent and agree with the literature, the underlying theory, the data collected, the constructs included, and the relationships developed, and it should minimize any possible misspecifications that can occur within the model. Then, the model and its degree of fit must also align with appropriately established fit indices. It should generally avoid second-order constructs, and it should include suitable validation processes.
This paper offers a checklist of requirements that can assist in the build of quality SEM studies.

SEM PROCEDURAL CONSIDERATIONS

Quantitative research engaging SEM often lacks in correct procedural development. In this study we present a basic set of requirements to assist in the deliverance of acceptable SEM solutions for business and for researchers.

Questionnaires

We begin with the questionnaires. Survey questions typically adapt previous literature-developed items - which in themselves are often modified to suite the current study, and studies often add loosely-related new items into their planned measurement model. Often items observed to work in another country, under differing situations, and often from different industries do not translate into a study's specific situation.
For example, selected items from various US studies may be adapted, applied, and assumed correct for a study in another country. This failure to triangulate for situational (time and place), industrial, and/or cultural/psychological differences can deliver a common method

bias. Such study weaknesses can initiate a set of Flaw 1 SEM problems, possibly resulting in low factor loadings within the measurement model.

Data correctness

Some studies do not report the treatment of missing values. In data entry and cleaning the SEM data sets should have no missing values, and only cases with less than 20% missing-at-random values [7] are acceptable for missing value replacement. Cases with 20% plus missing values warrant elimination from the SEM data set. This is Flaw 2.

The imputation of missing values is also often overlooked or unclear. The failure to report Little's MCAR [16] is Flaw 3 - as it confirms (or refutes) correct replacement as missing-completely-at-random.

Violation of normality assumptions

Flaw 4 is the failure to now investigate each item used for possible range weakness. Here, outliers or double peaking cases require rectification or elimination. The mean, SD, skew and kurtosis of cases used should be considered for each item. Strongly skewed data may negate the use of maximum likelihood (ML) – with principal axis factoring (PAF) being the preferred alternative in such cases.

Flaw 5 is where an item has both high kurtosis and high skew. Here the item should be considered for removal, as it only covers a small part of the possible Likert scale range, and so lacks discrimination.

Flaw 6 is where the item's mean is towards an extremity, and where twice its SD is beyond the Likert range (such as < 0.500 and > 7.499 for a seven point scale). In such situations this item should be considered for removal.

Factor analysis

Items engaged in a study are often adapted from previous research or from the literature. When conducting factor analysis for measurement models studies should apply confirmatory factor analysis (CFA). Flaw 7 is the failure to recognize the questionnaire study, built on literature developed questions, is actually CFA, and not exploratory factor analysis (EFA). CFA constructs are developed from within the literature and through the data set's collection framework. Each CFA construct captures the significant inter-group correlations of its embedded items (and the construct differ from all other constructs). When SEM is engaged (under ML) CFA is the default method assumed.

Sample size

Flaw 8 is using a survey that has a small number of respondents. This delivers sample size problems, and it affects the mathematics necessary for SEM calculations. The ratio of sample size to measured parameters is ideally 20:1 [23], but realistically can still work at 5:1 [3]. A sample size of 200 cases is also desirable, but models with: no latent constructs, all loadings fixed to 1, and strong correlations (or very simple models) can sometimes deliver reasonable model fit - even against only 75 to 100 cases [5]. Here $\chi 2$ is the most important fit index.

Calibration and validation

SEM studies often just report how their claimed model(s) fits against measures, and then do not explain how the claimed developed model(s) is validated, and has not emerged by chance. The SEM approach fits the model(s) to the data. Theoretically if the model is correct, then a second data set - capturing the same items and factors, should also fit - thus delivering a parsimonious and validated model. For example a large data set (for example 600) is randomly split into a calibration and validation data set. Theoretically, the model is first built with the calibration data set and should also work when run with the second (validation) data set. When a significant difference is observed between the calibration and validation model this indicates a possible model misspecification in the measurement model. This is rarely reported and so leads to Flaw 9.

Flaw 9 is non-sectioning of large data sets into confirmation (around 60%) and validation (around 40%) data sets (and still meeting minimum sample size requirements). This delivers a strong validation case [8]; [21]. The confirmation and validation data set should both typically hold around 20 plus cases per construct [4]; [7]; [9], and the validation data set ideally has 20 cases per construct [4]; [7]; [9].

Model validation can also be conducted for small data sets using a bootstrapping (between 200 to 2000 times) approach and reported with the Bollen-Stine p (> 0.05). For larger data sets, calibration and then validation modelling [20]; or in special cases [8], better establishes a valid model.

Construct development

SEM is influenced by the final construct items in the measurement model engaged. Flaw 10 is a failure to follow and report on a ML (or PAF) factor reduction process, or a failure to follow a congeneric construct process (and its items checking procedure). Here, one-at-a-time, poorly loading items, and cross-loading items, are individually considered for removal (or relocation onto another construct). Provided no theoretical repercussions exist, this generally delivers reduced discriminant validity issues.

From the factor reduction process each final construct consists of three (or four) strongest convergent/correlated items [4]. Often studies do not execute this step, and conveniently do not report on any factor reduction, or disclose the interaction (discriminant) effects between their construct items. However, in the final data reporting experienced SEM researchers/reviewers can pick such faults very quickly. For example, a low (< 0.6) or very high (> 0.95) Cronbach alpha (indicating the internal consistency of the constructs) suggests construct problems. In some cases an item on a construct dominates the overall construct loading - giving a high Cronbach alpha, but still holding an item bias.

In some cases average variance extracted (AVE > 0.5) is used to assess factor loading but this alone does not replace the Flaw 10 requirements for convergent/discriminant validity. Occasionally, only one or two items are deployed for a construct - but this is typically an unstable situation and undesirable for SEM.

So across the final CFA constructs engaged in SEM, we check construct's reliability - alpha > 0.7 (composite), AVE > 0.5 (convergent), and ensure construct discriminant meaning - with minimal cross-loading (maximum shared variance < AVE; average shared variance < AVE; inter-construct correlations < $\sqrt{}$(AVE) before moving towards a causal model.

Second-order constructs: part 1

Flaw 11 is the use of a second-order construct model to combine selected constructs (and their embedded correlated items) as per Figure 1 and Figure 2. Using data sets, free from the above Flaws 1 to 10, we have assessed higher order models and only once have established a second-order construct model that offered marginally better information to the alternate model without a second-order construct. It is preferable not include any second-order construct in SEM studies.

The Figure 1 and 2 model is often strengthened when more than one literature supported input construct is engaged, and when the maximum number of sequential paths remains under five (ideally two or three).

FIGURE 1: Second-order Measurement Model

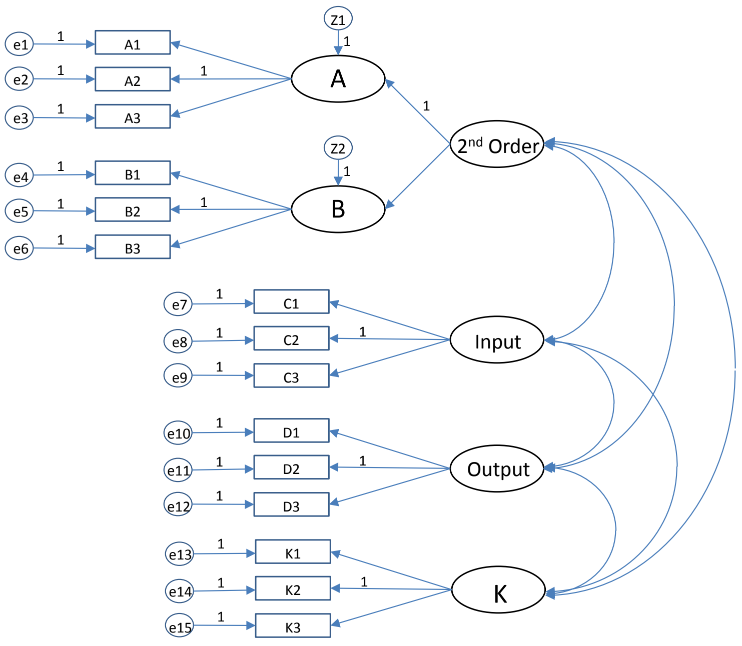
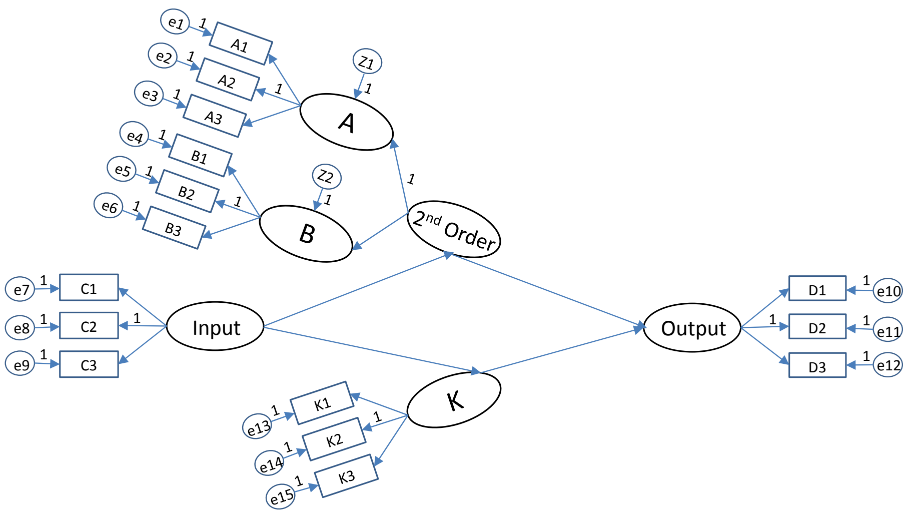


FIGURE 2: Second-order Structural Model



Applying modification indices (MIs) and standard residual covariance matrix (SRCM) values

Studies often report MI's as a main way to enhance the model's goodness of fit (Table 1). Flaw 12 uses MIs (> 4.0) to show discrepancies adjustments between the proposed and estimated model. Large MIs suggest where regression paths may be added.

To remove paths, standardized residual covariance matrix values (> 2.0) significantly reduce model fit, and indicate where paths could be eliminated to improve model fit. This is Flaw 13 However, where paths are to be added or removed, such changes are sometimes contrary to the literature – yet these additions are rarely justified, and they may also deviate from the initial theoretical model.

MIs between error terms should be noted - but not covaried, unless an exceptional or theoretical reason exists [12].
This use of MI's and SRCM values as the prime indicators when deciding whether to add or remove model paths between constructs is Flaws 12 and 13.

TABLE 1: SEM goodness of fit indices

Fit Index	Recognized Thresholds	Description
Absolute Fit Indices		
Chi-Square $\chi 2$	$\chi 2$ low compared to df ($p > 0.05$) 05 [4]; [7]	Closeness of model fit to data
Relative $\chi 2 = \chi 2/df$	$1 < \chi 2/df < 2$ [4]; [23]) $1 < \chi 2/df < 3$ [15]	Moves with size of data set
Root mean square error of approximation	RMSEA < 0.05 [4]; [7]	Known distribution & parsimonious. RMSEA < 0.05 = ex. fit Big models RMSEA < 0.08
Goodness of fit index	GFI > 0.95 [7]	GFI near 1. GFI > 0.95 desired Use cautiously
Absolute goodness of fit index	AGFI > 0.95 [7]	AGFI near 1. AGFI > 0.95 desired Adjusts for more parameters
GFI minus AGFI	(GFI-AGFI) < 0.06 [4]	If > 0.06 = model problems
Root mean square residual	RMR < 0.05 = good models [23]	Unstandardized, residual based Av. squared dif (sample covariances residuals & residuals estimated covariances)
Standardized root mean square residual	SRMR less than 0.08 [11]	When standardised is easier to compare.
Incremental Fit Indices		
Normed-fit index	NFI > 0.95 [7]	Assesses fit to baseline model - assumes no covariances between observed variables. Can overestimate fit when using small data sets
Tucker-Lewis index	TLI > 0.95 [4]; [7]; [17]; [20]	Non-normed, values just above 1 = OK. Good for small data sets or simulations
Comparative fit index	CFI > 0.95 [11]	Normed, good for small or larger data sets.

Fixing models

When MI's between inter-construct residual error items are large, model fit can be restricted. Here, a number of small refinements (or fixes) are possible - yet these seldom correctly applied. First fix is the check and elimination of any remaining outliers in the data set.
Second fix is to apply the appropriate fit measures. For small data sets [10]; [22], the model should tighten and deliver the appropriate fit key numbers (CFI > 0.95, TLI > 0.95, RMSEA < 0.05; GFI and AGFI both greater than 0.90 (ideally > 0.95); and (GFI – AGFI) > 0.06 [4]; [7].
Third fix is to examine the items on the construct that load poorly (and possibly should be on a different construct). Poorly loading items (such as 0.35 vs 0.85) can affect the construct's path strengths (regression weights), and should be considered for removal from the construct. Such removals can improve and sometimes solve invariance issues.
Fourth fix is to reduce inter-item interactions and invariance issues by adopting single item composites [6]. This approach builds one composite item from the set of construct items. It engages Munck's equations [19] to build the loading and the error term. This is very different from using an unstable single item construct as mentioned above in Flaw 10.
A fifth fix is to covary error terms, but as a general rule, this should not be done without precise reasoning [5]; [12], and ideally strong literature support.
If these fixes all fail to improve the model sufficiently, and the model still does not fit, then it is back to the drawing board.

Chi-square
Flaw 14 is the misunderstanding around the χ^2/df ratio [22]; [24]. When the final SEM model has $1 < \chi^2/df < 3$ (1 to 2 [23]; or 1 to 3 [15] and $p > 0.05$ [1] - then model fit occurs. Although Chi-square assumes multivariate normality it is generally a suitable fit measure provided the deviations are not extreme [18]. For very small data sets (< 75) Chi-square lacks discriminate power in distinguishing between good and weaker fitting models [14]. Hence, this is the key fit index for SEM – especially for small data sets. For models with above 400 cases other fit indices help [2]; [13].

Second-order constructs: part 2
Flaw 15 is the use of second- (higher-) order constructs to house related constructs and to relate these into models. Typically this approach does not add value to the model and often masks path, and progressive, interactions between constructs. Hence, such an inclusion must first be precisely established, and must be for a very specific purpose, and each factor included must be free of significant intra-construct and inter-construct interactions.
A second-order construct should only be included as an intermediate construct, and there should not be multiple (or just) second-order constructs in this intermediate stage of the model (as additional constraint requirements can be manipulated and so influence model performance). As per Flaw 11 if A and B in Figures 1 and 2 can be treated as individual constructs further intermediate paths and model knowledge are generally exposed.

DISCUSSION AND CONCLUSION

Although SEM remains a versatile statistical technique for research across disciplines its modelling is often delivered and explained using differing degrees of detail. Such inconsistencies often arise due to different journal/conference and reviewer requirements. This paper provides a set of 15 common Flaws as a checklist for SEM study evaluations to add clarity to SEM reporting. Beyond clarifying and assisting in overcoming these 15 SEM Flaws, this study also offers five additional ways to incrementally overcome SEM fit weaknesses.

REFERENCES

[1] Barrett, P. “Structural equation modelling: Adjudging model fit,” *Personality and Individual Differences*, 2007, 42(5), 815-24.

[2] Bentler, P.M., & Bonnet, D.C. ‘Significance tests and goodness of fit in the analysis of covariance structures,” *Psychological Bulletin*, 1980, 88(3), 588-606.

[3] Bentler, P.M., & Chou, C.P. “Practical issues in structural modeling,” *Sociological Methods & Research*, 1987, 16(1), 78-117.
[4] Cunningham, E. *A practical guild to structural equation modeling using AMOS*. Melbourne, Australia: Streams Statsline, 2008.

[5] Gerbing, D.W., & Anderson, J.C. “The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis,” *Multivariate Behavioral Research*, 1985, 20, 255−271.

[6] Grace, J.B., & Bollen, K.A. “Representing general theoretical concepts in structural equation models: the role of composite variables,”. *Environmental and Ecological Statistics*, 2008, 15(2), 191-213.

[7] Hair, J.F. Jr, Black, W.C., Babin, B.J., & Anderson, R.E. *Multivariate data analysis (7th ed.).*Upper Saddle River, NJ: Prentice-Hall, Inc., 2010.

[8] Hamilton, J. “Delivering service value networks: smart business,” *36th Western Decision Sciences Institute Annual Meeting*, 4-6 April, Denver, Colorado, 2007, 227-232.

[9] Hamilton, J.R., & Tee, S. “Smart utilization of tertiary instructional modes,” *Computers and Education*, 2010, 54(4), 1036-1053.

[10] Hooper, D., Coughlan, J., & Mullen, M.R. “Evaluating model fit: A synthesis of the structural equation modelling literature,” *7th European Conference on Research Methodology for Business and Management Studies*, Regent’s College, London, UK, 2008, 1-10.

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

[12] Jöreskog, K., & Long, J.S. “Introduction in testing Structural Equation Models,” In Kenneth A. Bollen and J. Scott Long, (eds.), Testing Structural Equation Model,. Newbury Park, CA: Sage, 1993.

[13] Jöreskog, K., & Sörbom, D. *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International Inc., 1993.

[14] Kenny, D.A., & McCoach, D.B. “Effect of the number of variables on measures of fit in structural equation modeling,” *Structural Equation Modeling*, 2003, 10(3), 333-51.

[15] Kline, R.B. *Principles and practice of structural equation modeling (2nd ed.).* New York, NY: The Guilford Press, 2005.

[16] Little, R.J. “A test of missing completely at random for multivariate data with missing values,” *Journal of the American Statistical Association*, 1988, 83(404), 1198-1202.

[17] McDonald, R.P., & Marsh, H.W. “Choosing a multivariate model: Noncentrality and goodness of fit,” *Psychological Bulleti*n, 1990, 107(2), 247.

[18] McIntosh, C. (2006). “Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett,” *Personality and Individual Differences*, 2007, 42(5), 859-67.

[19] Munck, I. *Model building in comparative education. Applications of LISREL method to cross-national survey.* Stockholm, Sweden: Almqvist & Wiksell International. 1979.

[20] Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W.R. “A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models,” *Journal of Business Research*, 2005, 58(7), 935-943.

[21] Standage, M., Duda, J., & Ntoumanis, N. "A test of self-determination theory in school physical education," *British Journal of Educational Psychology*, 2005, 75, 411-433.

[22] Tabachnick, B.G., & Fidell, L.S. *Using multivariate statistics (5th ed.).* New York: Allyn and Bacon, 2007.

[23] Tanaka, J.S. "How big is enough? Sample size and goodness of fit in structural equation models with latent variables," *Child Development*, 1987, 58, 134-146.

[24] Wheaton, B., Muthen, B., Alwin, D.F., & Summers, G. "Assessing reliability and stability in panel models," *Sociological Methodology*, 1977, 8(1), 84-136.