This is a repository copy of *Ten basic questions about structural equations modeling you should know the answers to – But perhaps you don't*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/163820/

Version: Accepted Version

Article:

Ten basic questions about Structural Equations Modeling you should know the answers to – but perhaps you don't.

Vasileios Davvetas

Lecturer (Assistant Professor) of Marketing

Marketing Division, Leeds University Business School, UK

v.davvetas@leeds.ac.uk

Adamantios Diamantopoulos

Chaired Professor of Marketing

Department of International Marketing, University of Vienna, Austria

adamantios.diamantopoulos@univie.ac.at

Ghasem Zaefarian

Associate Professor of Marketing

Marketing Division, Leeds University Business School, UK

g.zaefarian@leeds.ac.uk

Christina Sichtmann

Associate Professor of Marketing

Department of International Marketing, University of Vienna, Austria

christina.sichtmann@univie.ac.at

July 2020

Abstract

Structural Equations Modeling (SEM) has enjoyed increased popularity as an analytical method among Industrial Marketing Management (IMM) authors over the last years. Despite such popularity, many authors fail to understand the basic principles of the method and reviewers are frequently confronted with manuscripts suffering from erroneous applications, insufficient reporting and questionable interpretation of SEM-based findings. Addressing this issue, the present article presents – in non-technical language – the most basic concepts related to SEM, resolves common misconceptions about the method's application and provides hands-on advice to IMM authors and reviewers dealing with SEM-based manuscripts. Structured along ten fundamental questions, the article covers issues related to (1) latent variables and their scaling, (2) types of parameters in SEM, (3) unstandardized and standardized estimates, (4) model identification, (5) model constraints, (6) model fit, (7) independence and saturated models, (8) modification indices, (9) nested models, and (10) equivalent models. After illustrating these concepts with the use of examples, the article concludes with a list of guidelines addressed both to IMM authors crafting manuscripts using SEM and the peers reviewing them.

Keywords: structural equations modeling, confirmatory factor analysis, survey research

1. Introduction

Structural Equations Modeling (henceforth SEM) is one of the most commonly used analytical methods in social sciences ranging from psychology and communication studies to international business and marketing (Holbert & Stephenson, 2002; Hult et al., 2006; Kumar, Sharma, & Gupta, 2017; MacCallum & Austin, 2000). Due to the extensive use of data collected through surveys in the industrial marketing field, SEM has been increasingly popular among *Industrial Marketing Management (IMM)* authors, too. Over the last 10 years, hundreds of publications in the journal have employed some analytical process related to SEM such as confirmatory factor analysis, path analysis, or scale development (Figure 1).

Insert Figure 1 about here

Despite such popularity, though, it is rather common to observe mistakes in the application of the method that range from minor issues such as incomplete or "selective" reporting of needed SEM statistics, to more severe errors such as flawed interpretations of SEM-obtained results in manuscripts submitted to IMM. These shortcomings result in the rejection of many submissions as they threaten key cornerstones of the scientific process such as the use of appropriate measurement instruments and the statistically sound testing of theoretical hypotheses. The underlying reasons for such shortcomings lie – more often than not – in researchers not having a *clear understanding* of some (very) basic principles of SEM but still applying the method due to the availability of user-friendly SEM software.[1] Indeed, there seems to be a strong need for the IMM community of authors and reviewers to

---

[1] Particularly the availability of software with graphical interfaces (e.g., AMOS) has been both a blessing and a curse for SEM. A blessing because it alleviates the need for complex programming and a curse because modeling is often reduced to "drawing" without fully understanding *what* is drawn and *how* to (correctly) interpret it.

familiarize themselves with key principles of SEM when developing and reviewing research manuscripts that use the method.

Against this background, the objective of the present paper is to provide correct answers to a list of the most fundamental questions relating to SEM, answers that *all* members of the IMM community should be aware of.[2] Our intention is not to provide an advanced technical discussion of SEM statistical metrics (e.g., alternative fit indices), estimation methods (e.g., maximum likelihood), or cases of special applications (e.g., estimation of latent means). Neither do we focus on data-related issues such sample size determination, distributional assumptions violations, or treatment of missing data. There is an extensive list of textbooks (e.g., see Bollen, 1989; Kline, 2015; Schumacker & Lomax, 2016) and specialized method journals (e.g., *Structural Equation Modeling; Multivariate Behavioral Research; Organizational Research Methods*) covering these topics in an excellent manner. Instead, we follow a "stick to the basics" approach, focusing on communicating – to a not necessarily expert audience – some fundamental concepts on which the healthy application of SEM relies.

We organize these concepts in the form of ten specific questions based on our personal experiences in applying SEM in our own research, reviewing SEM-based articles in a variety of marketing and management journals, and teaching SEM courses to research students and junior faculty. The answer to each question addresses a core concept in SEM that is relevant and applicable *irrespective of the specific model or data involved.* These core concepts are thus of the "must know" variety just as the location (and function!) of the accelerator and brake pedals are "must know" elements for safely driving a car, *any* car. Understanding these

---

[2] In this article, we focus exclusively on covariance-based SEM as implemented in software packages such as LISREL, EQS, AMOS or Mplus. Readers interested in partial least squares (PLS) modeling are referred to Hair, Hult, Ringle, & Sarstedt (2016), Sarstedt, Ringle, & Hair (2017) and references given therein.

concepts should help researchers make the most out of SEM in their empirical applications and remove any uncertainty, guesswork, or confusion associated with the procedure.

The rest of the paper is organized as follows: We present each of the ten questions in a standalone section where we discuss key issues that help answer the specific question. Subsequently, we provide an illustration of these issues using a simple model to showcase the application of the method and offer a good example of the method's use and reporting. Finally, we develop a table offering guidance to authors by summarising the most important "do's and dont's" they need to bear in mind when drafting SEM-based manuscripts. We also provide appropriate guidelines to reviewers by directing their attention to important questions they should pose to authors during the review process to assess the appropriate use of SEM and the validity of the resulting findings.

2. Ten Basic Questions (and Answers!) about SEM

SEM, as an analytical method, can be simply understood as a combination of factor analysis and multiple regression modeling. The factor analysis element of a SEM model is focused on assessing the appropriateness of the variables used in the model, while the multiple regression element is focused on estimating the hypothesized effects of some variables on others. SEM is particularly useful when researchers deal with data obtained through questioning respondents via primary data collection methods such as surveys and experiments. Because such data cannot be readily obtained through secondary sources and often constitute the only way to test theoretical hypotheses of interest, SEM has emerged as a valuable analytical tool in management and related domains where the effects of certain psychological/organizational/strategic concepts are paramount to theory building. Self-report data do not represent "perfectly measured information" lacking measurement errors and are

easily affected by issues that potentially hurt the validity and reliability (e.g. respondent fatigue, social desirability biases, common method variance, etc.) of corresponding findings if modeled through other analytical tools. SEM is particularly effective with dealing with this kind of issues, thus helping reach theoretically and empirically sound conclusions.

In light of the above, SEM emerges as a necessary modeling tool in the following research contexts, among others. First, being able to assess the covariance structure of variables through confirmatory factor analysis in a more stringent manner than exploratory factor analysis approaches. SEM is particularly helpful for testing the measurement instruments' psychometric properties and isolating measurement errors that would hurt the tests of theoretical propositions. Second, SEM is particularly useful when researchers are interested in developing novel measurement scales for constructs and thus in need to assess the validity, reliability and predictability of such scales for future applications. Finally, unlike typical regression models trying to estimate effects of a set of independent variables on a single criterion variable, SEM offers the ability to simultaneously estimate substantially more complex model structures that involve variables operating simultaneously as both causes and outcomes of other variables in the model. This enables the estimation of both direct and indirect effects among a set of variables of interest.

2.1. Question 1: What are latent variables and how should they be scaled?

Unlike variables that can be directly measured using objective data such as revenues, profits, costs or number of customers, researchers are often interested in measuring theoretical constructs that cannot be effectively quantified using secondary, observable information – such as a retailer's satisfaction with a supplier or the level of trust between two business partners. These unobservable variables are called *latent variables* and represent constructs which are typically measured through survey instruments administered to managers,

employees or other key organizational informants.[3] To quantify a latent variable, researchers typically rely on those informants' scores on several items (indicators) intended to capture the construct of interest. Such items are commonly called *manifest (or observed) variables* because, unlike latent variables, they *can* be directly observed (usually using reported scores on Likert-type or other rating scale formats). Although manifest variables are necessary for measuring the latent variables of interest, their psychometric nature makes them prone to issues of validity and reliability such as measurement errors resulting from respondents' response biases or measurement instrument characteristics. One of the key benefits of SEM is the fact that relationships between latent and manifest variables can be formally specified (typically in a linear form) and measurement error explicitly accounted for. Generally, constructs can be measured either in a reflective manner (i.e., through specifying the manifest variables as causal manifestations of the latent variable and whose variability is predominantly explained by the latent variable) or in a formative manner (i.e., through specifying the manifest variables as elements that causally form the latent variable and contribute to its variance).

Going from respondents' scores on a set of observed (manifest) variables to the quantification of an unobserved (latent) variable, though, requires the specification of a scale format for the latent variable. This is necessary because, being unobserved, latent variables do not have a "natural" unit of measurement (e.g., what are the units of measurement of "trust" or "relationship quality"?). The process of assigning a unit of measurement to a latent variable is called *latent variable scaling*. This is typically achieved in one of two ways. First, one can simply *standardize* the latent variable by setting its variance to 1 (some SEM software, such as LISREL do this by default). Second, one can select one of the manifest

---

[3] The terms "constructs" and "latent variables" are often used interchangeably in literature. Strictly speaking, however, latent variables are *representations* of constructs in SEM. Thus, in the case of multidimensional constructs, several latent variables may be needed to formally represent them in a model.

variables as the *reference* or *scaling indicator* and set the value of its loading (i.e., the coefficient capturing the association between the latent variable and the indicator) equal to 1. Importantly, using a reference indicator and fixing its loading to 1 does *not* set the latent variable equal to the observed indicator. What it does, is assign the units of measurement of the reference indicator to the latent variable; this means that the variance of the latent variable will be estimated in the units of measurement of the reference indicator. Any one manifest variable could be used as a scaling indicator – without affecting the conceptual meaning of the construct, the estimation of its effects, or the fit of the model.[4]

Quite often authors fail to mention in their manuscripts what their scaling indicators are; sometimes, authors do not even report *any* of the indicators used to measure their constructs (referring instead to past papers that used the same operationalization); and sometimes, the indicators themselves are reported without, however, any relevant psychometric information (e.g., the loadings and error variances of the manifest variables – see Question 2 below). Such poor reporting practices prevent readers from properly judging the operationalization of the latent variables used to represent the constructs of interest in the model. Thus, authors are strongly advised to provide a *complete list* of the items used to measure every latent variable in their models and highlight which of these items serve as scaling indicators.

To illustrate the above points, imagine that a researcher wants to test a model depicting how the stereotype a client has about the supplier's staff impacts the client's satisfaction with the supplier and the willingness to repurchase from this supplier. Let us assume that, after reviewing related literature, the researcher proposed a model where two fundamental

---

[4] Note that this holds for reflectively-measured constructs *only*. For formatively-measured constructs, researchers should consider alternative options for scaling the latent variable (for details, see Diamantopoulos, 2011; Diamantopoulos & Riefler, 2011).

dimensions of the supplier staff stereotype (competence and warmth) affect satisfaction with the supplier and, through it, repurchase intent (Figure 2).

Insert Figure 2 about here

The model includes two *exogenous* latent variables[5], namely supplier staff competence (ξ_1, capturing supplier staff ability to competently satisfy the client's business needs – measured with five items, x_1-x_5) and supplier staff warmth (ξ_2, capturing the supplier staff's positive intent toward the client – also measured with five items, x_6-x_{10}). The model also includes two *endogenous* latent variables[6]: satisfaction with the supplier (η_1, measured through three items y_1-y_3, completed by the manager handling the supplier, the team working with the supplier, and the CEO of the client) and repurchase intent (η_2, also measured by three items y_4-y_6 from the aforementioned informants). In short, the model in Figure 2 contains four latent variables and a total of 16 manifest variables (indicators).[7] In this example, if item x_1 = "the supplier's staff is competent" is chosen as the scaling indicator for the supplier staff competence construct, the researcher would set $\lambda_{11} = 1$; if item x_6 = " the supplier's staff is warm" is selected as the scaling indicator for the supplier staff warmth construct, the researcher would set $\lambda_{62} = 1$, and so on.

Note that the number of manifest variables in a model determines the *measurement model relations,* that is, the equations linking the latent variables to their measures; in Figure 2 there are a total of 16 such equations, of which ten relate to the measures of the exogenous latent variables (i.e., $x_1 = \lambda_{11}\xi_1 + \delta_1$ through to $x_{10} = \lambda_{1,10}\xi_2 + \delta_{10}$) and six to the measures of the

---

[5] Exogenous variables act *always* as independent (predictor) variables and never have error terms.
[6] Endogenous latent variables act *always* as dependent (criterion) variables and always have error terms; however, they can also act as independent variables impacting other endogenous variables (e.g., see η_1 in Figure 2).
[7] As an aside, note that η_1 is modeled as a *mediator* of the relationship between ξ_1 and η_2 as well as of the relationship between ξ_2 and η_2.

endogenous latent variables (i.e., $y_1 = \lambda_{11}\eta_1 + \varepsilon_1$ through to $y_6 = \lambda_{62}\eta_2 + \varepsilon_6)$. The number of

endogenous latent variables determines the *structural model relations*, that is, the equations

linking the latent variables to one another. In Figure 2, there are two such equations one

relating to the satisfaction with the supplier ($\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1$) and one to repurchase

intent ($\eta_2 = \beta_{21}\eta_1 + \zeta_2$). Finally, the number of exogenous latent variables determines the

number of *non-directional relationships* (covariances) in the model. In Figure 2, there is only

one such covariance (i.e., $\varphi_{12} = COV(\xi_1\xi_2)$). For an overview of key SEM terms, their

definitions and the corresponding notation, see Table 1.

Insert Table 1 about here

2.2. Question 2: What kind of parameters are we interested in when estimating in SEM models?

In SEM models, researchers are interested in the estimation of several types of parameters.

First, researchers are interested in the *loadings* (λ's) and *error variances* (VAR(δ)'s and

VAR(ε)'s) of the manifest variables; the loadings capture the association between the

manifest variables and the latent variable, while the error variances capture the remaining

(residual) variation of the manifest variables after the influence of the latent variable has been

accounted for. These parameters are referred to as *measurement model parameters* and offer

a test of whether the chosen manifest variables are valid and reliable measures of the latent

variable (and thus whether using them can be safe when testing for hypothesized

relationships among latent variables). From the loading and associated error variance, one

also obtains the *squared multiple correlation* (SMC) for each manifest variable. This is

essentially an R^2 value that shows the proportion of variance in the manifest variable

attributable to the underlying latent variable (e.g., a SMC value of .80 indicates that 80% of

the variance in the manifest variable is due to the latent variable to which it is assigned, while the remainder 20% is due to measurement error). Researchers are also often interested in the *covariances/correlations* between latent variables which are important for validity tests (e.g. discriminant validity assessment).

Second, researchers are interested in *structural model parameters*, that is, the *path coefficients* (γ's and β's) capturing the hypothesized non-zero directional effects between two latent variables in the model as well as the *structural error variances* (VAR(ζ's)), from which *squared multiple correlations* can be calculated. The latter capture the proportion of variance in an endogenous latent variable that is explained by the independent latent variables that influence it (i.e., similar again to the R^2 statistic obtained in conventional linear regression).

Irrespective of the software package used, SEM estimation generates three values: (a) the parameter estimate (e.g., loading or path coefficient), (b) its corresponding standard error (i.e., the standard deviation of the sampling distribution of the parameter), and (c) its *t*- or *z*-statistic which determines the significance level of the parameter (and is nothing more than the simple fraction of the (unstandardized) parameter estimate over its corresponding standard error).

Measurement and structural model parameters are of equal importance when testing theoretical relationships and both should have acceptable values for any findings to carry meaning. On the one hand, poor measurement model parameters (e.g., low SMCs) imply weak measurement of the latent variables, rendering even significant structural parameters questionable or even misleading. On the other hand, although structural model parameters are tests of hypothesized theoretical relationships and researchers want them to have statistically significant values, they can still be informative by being non-significant (e.g., by suggesting that a hypothesized relationship does not receive empirical support by the data).

Quite often, authors pay less attention to measurement model parameters in their manuscripts because they rely on scales that have been used in prior literature and/or because they perceive measurement parameters to be irrelevant for hypothesis testing. However, merely *assuming* that a measurement instrument successfully used in the past is necessarily psychometrically sound when applied to one's *own* data may prove questionable. Appropriate estimation and discussion of measurement model parameters is a prerequisite for meaningful statistical inferences about structural model parameters. Similar to how any body temperature figure is uninformative (if not dangerous) when it is recorded with a broken thermometer, any structural parameter estimate is useless unless it is obtained using psychometrically-sound measurement instruments. Thus, researchers should *always* include a detailed list of the measurement instruments they use accompanied with the corresponding psychometric properties and present the results of their measurement model *before* proceeding with the results of hypothesis testing.

An example of a measurement model using the same latent variables as in Figure 2 is shown in Figure 3. It is worth noting that although both structural and measurement models include the same latent and manifest variable, in the measurement model, every latent construct is allowed to correlate freely with *all* other latent variables in the model; these covariances (or correlations if we assume standardization – see Question 3) are denoted by the φ's and are captured by the two-headed arrows in Figure 3. In contrast, in the structural model of Figure 2, only the directional paths (i.e., γ's and β's linking different latent variables that are *theoretically* expected to be related are included; the only covariances/correlations specified are those between the exogenous variables (i.e., φ_{12})).

Insert Figure 3 about here

12

2.3. Question 3: What are unstandardized and standardized estimates?

SEM software usually reports estimated parameters in both *unstandardized* (raw) and *standardized* form. Researchers are often unclear regarding the difference between the two and which they should report in their manuscripts. To illustrate, take the example of a path coefficient between an independent latent variable (satisfaction with supplier X) and a dependent latent variable (repurchase intent from supplier X) and imagine that its estimated unstandardized value is $\beta_{21} = 0.5$ (also assume that it is statistically significant). This estimate can be interpreted as: an increase of one unit in the scale measuring supplier satisfaction is associated with a 0.5 unit increase in the scale measuring repurchase intent, all other variables held constant. The word of interest here is *unit of measurement*, as different units of measurement lead to different implications for the practical significance and interpretability of the observed effect.

Consider three scenarios where the 0.5 path estimate is obtained. *Scenario 1*: if satisfaction and repurchase intent are both measured on a 0-100 scale, then the effect is rather weak bearing in mind the scale range. *Scenario 2*: if satisfaction and repurchase intent are both measured on a 1-5 Likert scale, the effect is rather strong. *Scenario 3*: if satisfaction is measured on a 0-100 scale and repurchase intent is measured as the "logged difference of the revenues from this customer over the last two years", interpreting the effect and its size could be rather troublesome for most.

To circumvent these problems, researchers often turn to standardized path coefficients which show the change in the dependent variable associated with an increase/decrease of one standard deviation in an independent variable, all other variables held constant.[8] Apart from their easier interpretability when different scale formats for dependent and independent

---

[8] A standardized path coefficient equals the value of the unstandardized parameter times the ratio of the standard deviations of the independent to the dependent variable.

variables are used, standardized coefficients enable *effect size* comparisons among different independent variables, even when the latter are measured in totally different units. Note that for non-directional paths (i.e., covariances), standardized estimates are simply the (bivariate) *correlations* between the variables involved.

Standardized parameters are particularly useful when researchers use numerical figures to measure directly observable dependent variables (e.g., sales figures). As these variables are often measured in thousands or even larger units, it is not uncommon to obtain unstandardized coefficients whose numerical values are statistically significant despite having nominally zero values. In such cases, it makes sense to report standardized coefficients that paint a more meaningful picture of the impact of the independent variable.

Overall, we advise researchers to report both standardized and unstandardized parameters when presenting their results or, at the very least, to *explicitly* state what type of parameters are reported in the text or in relevant tables.

2.4. Question 4: What is model identification?

To estimate a model in SEM, one uses existing pieces of information (e.g., data inputs from surveys or experimental instruments) to generate new pieces of information (i.e., estimates of theoretically important parameters). In SEM, there are two types of information pieces provided in a dataset: the *covariances* between any pair of manifest variables and the *variances* of the latter (i.e., the observed variables' covariances with themselves). In any structural equations model, the number of available pieces of information is given by the formula: $s = k \times (k + 1) / 2$, where k is the number of manifest variables included in the model (regardless of which latent variable they are assigned to). Turning to the pieces of information a researcher needs to estimate, their number depends on the model setup (i.e., how many free and constrained parameters exist in the model). As further discussed in

Question 5, a *model constraint* decreases the pieces of information needed to be estimated by fixing a parameter's value (e.g., a zero path) or by requiring it to have the same value with another parameter (in which case one obtains one estimated value for two or more parameters).

Model identification refers to whether the researcher has enough pieces of information to obtain unique estimates of the parameters to be estimated (known as *free parameters*). The difference between available pieces of information (i.e., variances and covariances of the observed variables) and free to be estimated parameters represents what is known as the model's *degrees of freedom*. For a structural equation model to be identified, the number of parameters that need to be estimated should *always* be less than or equal to the number of unique pieces of information provided by the data; in other words, the model should have non-negative degrees of freedom. Otherwise, the model is *under-identified*, that is, no unique parameter estimates can be obtained (and thus no testing of hypotheses is feasible). If the pieces of information provided by the data exactly equals the number of parameters to be estimated, the model becomes *just-identified*, meaning that unique parameter estimates are provided but the overall *model fit* (see Question 6) cannot be tested. Finally, if the available pieces of information exceed the number of parameters to be estimated, the model is *over-identified*, that is, one can both obtain more than one set of estimates of the model parameters and use these additional estimates to test the model. For this reason, researchers are urged to develop over-identified models by ensuring that the following relationship holds: $t < s$, where t = number of parameters to be estimated and s = the total number of (unique) variances and covariances among the observed variables.

The importance of model identification is illustrated in Figure 4. Imagine for a moment that the researcher's goal is not to test the overall model earlier presented in Figure 2 but instead to simply test the supplier staff competence scale on its own merit. Let us further

assume that the researcher is considering three alternative scales: one with two items, one with three items and one with four items, and that the latent variable (ξ_1) has been scaled by standardizing it (i.e., fixing its variance to 1).

Insert Figure 4 about here

In the case of the two-item scale, the researcher has three available pieces of information from the data (i.e., the variances of the two items x_1 and x_2, and their covariance) but needs to estimate four measurement parameters (i.e. the two loadings λ_{11} and λ_{21} and the variances of the corresponding errors δ_1 and δ_2). In this case, the model cannot be estimated because it has negative degrees of freedom. In the case of the three-item scale, the researcher has as many available pieces of information (i.e., the variances of the three items x_1, x_2 and x_3, and the three covariances between the items) as those s/he needs to estimate (i.e., three loadings λ_{11}, λ_{21} and λ_{31} plus the three variances of the error terms δ_1, δ_2 and δ_3). This makes the model just-identified, that is, all parameters can be estimated but the model cannot be tested as the degrees of freedom are zero. Finally, in the case of the four-item scale, the researcher has ten pieces of information (i.e., the variances of the four items x_1, x_2, x_3 and x_4, and the six covariances among the items). The last case leads to an over-identified model with two degrees of freedom which allow testing the model's fit.

The above example illustrates that whether a model will be over-, just-, or under-identified is not something one learns *after* conducting the analysis. Instead, it is something that a researcher can (and should) check *before* collecting any data to avoid unpleasant surprises (which would be preventable if identification issues had been considered in advance). It also illustrates that scales with a limited number of indicators might create problems of identification when estimating measurement models. In general, four options

exist to overcome identification problems: (1) setting some parameters to fixed values (e.g., by removing a structural path), (2) setting parameters equal to each other (e.g., specifying two loadings to be the same), (3) introducing additional information (e.g., adding an indicator) or (4) embedding the model within a bigger model (e.g., while the two-indicator measurement model in Figure 4 is under-identified, a model with *two* latent variables with two indicators each, is over-identified with 1 degree of freedom).

2.5. Question 5: What are constraints in SEM?

There are two types of parameters in SEM: free parameters and constrained parameters. *Free parameters* are those that researchers want to estimate in order to assess the quality of their measures and test their hypotheses; typical free parameters are the loadings and error variances of manifest variables and the hypothesized paths between two latent variables. *Constrained parameters*, on the other hand, are parameters that are specified to have either a fix numerical value (e.g., zero) or parameters that have the same value with some other parameter in the model (e.g., two manifest variable loadings set equal).

Recalling the discussion of Question 4, it is evident that constraints affect model identification. As constrained parameters do not need to be estimated, they increase the model's degrees of freedom. More specifically, introducing a constraint will *always* lead to a deterioration of model fit (even if insignificant) while relaxing a constraint will *always* lead to an improvement in model fit (even if insignificant). Furthermore, models that are under- or just-identified can become over-identified through the introduction of constrained parameters; and over-identified models can turn to under- or just-identified ones by relaxing constrained parameters. Although introducing arbitrary constraints in a model purely to achieve model identification is generally ill-advised and alternative options should be considered to solve identification issues (e.g., avoiding single item measures), doing so might

be justified in some cases (e.g., when one can make a reasonable assumption that all items of a reflectively-measured latent variable carry equal loadings).

Apart from model identification issues, model constraints are important for theory testing purposes and should be assessed with care because they can have serious implications for the theoretical soundness and the practical relevance of obtained results. Unfortunately, it is common practice for researchers to either introduce constraints that should not be included in the model based on theoretical arguments or freeing constraints by allowing the estimation of relationships that should conceptually be set to zero. Typical examples of the former case include setting the covariance between two exogenous variables to zero (which can be harmful as exogenous variables should almost always be allowed to correlate in a model) or setting measurement error variances equal to zero (which implies lack of measurement error in manifest indicators). Typical examples of the latter case include allowing the free estimation of the covariance between the structural error terms of two endogenous variables (which should not be allowed unless based on theoretically-grounded arguments of the presence of a *common* omitted predictor variable) or allowing the free estimation of error covariances between manifest indicators measuring the same latent variable (which implies the undesirable presence of other sources of common variation beyond the latent variable, and thus, questions regarding the validity of the construct's measurement). All these actions lead to artificial inflation or deflation of model fit indices in ways which do not enhance the theoretical value of the model.

Researchers typically focus more on free parameters as these correspond to theoretical relationships to be tested. In doing so, they often miss constrained parameters that become part of the model "by default" (i.e., without being consciously/intentionally constrained by the researcher). To illustrate this point, Figure 5 shows that, in the illustrated model earlier presented in Figure 2, there are two parameters (γ_{21} and γ_{22}) – the direct paths from supplier

staff stereotype dimensions to repurchase intention – that have been set to a fixed value (namely, zero). These constraints imply that the authors do *not* theoretically expect any effect of supplier staff competence or warmth on repurchase intent that is unaccounted by supplier satisfaction; in other words, supplier satisfaction is hypothesized to *fully* mediate the impact of competence and warmth on repurchase intent. These constraints thus have *theoretical* relevance and can be potentially challenged by reviewers. Specifically, if one can make a theoretical case about the inclusion of these direct paths in the model, then these constrained parameters should be set free and tested along with the other free parameters. Moreover, their contribution to significantly improving model fit should be noted. Thus, authors who choose to exclude these paths from estimation (i.e., assume they are zero as in Figure 5), should have a compelling *theoretical* argument at hand for doing so and make this argument available to reviewers when describing their model.

Insert Figure 5 about here

Overall, we advise authors to (1) be fully aware of the constraints included in their models, (2) be prepared to theoretically defend them, and (3) not relax model constraints using favorable changes in model fit as the only justification for doing so.

2.6. Question 6: What is model fit?

One of the main concerns of researchers when using SEM is whether their model has a good *overall fit*. Although most authors know that they need to report model fit statistics and reviewers require them to reach "make or break" decisions on manuscripts, the concept of fit is perhaps the most misunderstood concept in SEM. While most researchers know that, in

broad terms, model fit captures the extent to which a hypothesized model is "in harmony" with the empirical data, few understand the *exact* notion of fit in the context of SEM.

Model fit captures the degree to which the data used to estimate the model (i.e., the sample covariance matrix $\mathbf{S}$) resembles the form that the data *should* have had if the hypothesized model were true in reality (i.e., the implied covariance matrix $\widehat{\mathbf{\Sigma}}$)[9]. Thus, model fit involves the comparison of two covariance matrices: the covariance matrix of the manifest variables based on the actual data and the covariance matrix of the manifest variables implied by the model. The bigger the resemblance or congruence between those two covariance matrices, the better the overall fit of the model. The formal test of this congruence (i.e., that $\mathbf{S} = \widehat{\mathbf{\Sigma}}$) is provided by a χ^2 statistic with degrees of freedom equal to the degrees of freedom in the hypothesized model.

Unfortunately, relying only the χ^2 test results to assess model fit is problematic as the test is affected by departures from multivariate normality, is sensitive to sample size, and also assumes that the model fits perfectly in the population. As a result, over the years, a long list of additional *fit indices* have been proposed in the literature along with proposed threshold values for judging the acceptability of the model.[10] While this is not the place to discuss the merits and shortcomings of different fit indices, researchers should always use *multiple* indices when evaluating overall model fit and avoid "cherry-picking" fit statistics to paint a more favourable picture of their models. Thus, from an author's perspective, the inclusion or exclusion of any fit index should not be made on the grounds of impression management but for purposes of striking a balance between parsimony and transparency in reporting.

---

[9] Formally, the implied covariance matrix $\widehat{\mathbf{\Sigma}}$ is the covariance matrix that would be obtained if values of the fixed parameters and estimates of the free parameters were substituted into the measurement and structural equations which were then used to generate a covariance matrix.

[10] Amongst the most popular fit indices are the RMSEA (Root Mean Squared Error of Approximation), the NNFI (Non-Normed Fit Index – also known as the Tucker-Lewis Index (TLI)), the CFI (Comparative Fit Index) and the RMSR (Root Mean Squared Residual). For a more detailed discussion of alternative fit indices and acceptable cut-off values per fit index, see Niemand & Mai (2018), Iacobucci (2010) and Steenkamp & Van Trijp (1991).

Conversely, reviewers should avoid overreliance on a couple of "popular" fit statistics (e.g., RMSEA or CFI), familiarize themselves with the strengths/weaknesses of particular fit indices in particular conditions (e.g., how sample size or number of estimated parameters can inflate or deflate a certain fit index) and appreciate that a value in one fit index that falls slightly short of conventional thresholds does not render a model automatically invalid if other fit indices show satisfactory values.

The above discussion has focused on the assessment of the overall fit of the model i.e. on *global fit*. However, in addition to global fit, researchers are (and should be) interested in individual parameter estimates and their implications for measurement quality and the validity of their theoretical predictions (known as *local fit*). It is important to appreciate, in this context, that a model with an acceptable global fit can have a fair share of "bad" local fit indices (e.g., non-significant path estimates) and, conversely, a model with a poor global fit can have most (if not all) structural path coefficients significant. Thus, for a model to be supportive of one's theory, both local and global fit need to be satisfactory. Importantly, global and local fit should be assessed both for the measurement model and for the structural model and researchers should appreciate the fact that sources of poor fit observed during measurement model assessment will inevitably spill over to structural model fit.

2.7. Question 7: What is an independence and a saturated model?

In a typical structural equations model, there are some parameters that are free to be estimated (e.g., hypothesized paths) and some parameters that are constrained to some fix value (e.g., zero covariances among error terms). When assessing overall fit in SEM, one sometimes needs a reference model. Two widely used reference models are the *independence model* and the *saturated model*. An independence model is one that does not allow any covariances among the observed variables, implying that every variable is orthogonal to all

the others. Such a model has the maximum degrees of freedom and the highest parsimony because most parameters are fixed to zero and thus almost nothing needs to be estimated (apart from the variances of the observed variables). A saturated model, on the other hand, is one where all observed variables are allowed to covary with each other and all available pieces of information are used to estimate the model parameters, making the model just-identified (i.e., having zero degrees of freedom). In other words, an independence model is a model full of constrained parameters while a saturated model is a model full of free parameters. Thus, any model with *some* free and *some* constrained parameters (i.e., the typical case of a model researchers try to estimate) can be seen as a more constrained version of the saturated model and a less constrained version of the independence model. If the independence and the saturated models are seen as the two ends of a road, each step from the independence model toward the saturated model requires setting a parameter free to be estimated at the exchange of one degree of freedom. In this sense, a degree of freedom can be understood as the price a researcher must pay to obtain an estimate of a parameter previously constrained to a fixed value. Conversely, every time a free parameter is constrained (i.e., set to a fixed value or equal to another free parameter), there is a gain of one degree of freedom.

More often than not, researchers do not explicitly mention in their manuscripts the relevant independence and saturated models or their corresponding fit statistics. Although such reporting is not directly needed to assess how good their proposed model is in terms of hypothesis testing and theory development, independence and saturated models are important because many of the global fit indices commonly used rely on comparisons between a hypothesized model and the relevant independence/saturated models. For instance, one of the most widely reported fit index in SEM is the Comparative Fit Index (CFI) which is calculated by comparing a candidate model with an independence model. Although we would not advise extensive reporting on independence and saturated models, it is still important for researchers

to understand the essence of those models when assessing the overall fit of their hypothesized model.

2.8. Question 8: What is a modification index?

When researchers estimate SEM models, they are often interested in *model modification*, that is, post-estimation alterations in the model that substantially improve overall model fit. Such alterations include the estimation of a new model by freeing one or more constrained parameters (see Question 5). When one previously constrained parameter is set free to be estimated, one degree of freedom is lost (because one needs to estimate one more parameter with the same available pieces of information) and the model fit (always) improves (because one "coaxes" the sample covariance matrix and the implied covariance matrix to become more similar).

A *modification index* is a statistic that shows the minimum improvement in model fit (in the form of chi-square reduction) that would be achieved if the model was re-estimated with a specific, previously constrained parameter set free to be estimated, while the rest of the model remains unchanged (modification indices are usually accompanied by the estimate of the expected value of that parameter as well). The presence of many and sizable modification indices implies that the original model is misspecified, meaning that many of the relationships that should have been included in the model as free parameters were not (and were thus wrongly specified as zero relationships or as values equal to another freely estimated parameter).

SEM software automatically produces lists of the larger modification indices in its effort to help researchers identify which model changes would make the model fit significantly better on the available data. Unfortunately, researchers have often been using these indices in an inappropriate manner, resulting in models with acceptable global fit but

lacking in theoretical soundness. Given that (1) modification indices provide specific

information on *which* parameter(s) should be set free, and (2) setting any parameter free

always leads to a better fit, it is, unfortunately, rather common for researchers to continuously

re-specify their original models post-hoc by allowing the free estimation of paths they have

previously constrained until a favorable global fit is achieved. This "end justifies the means"

approach, though, violates the principles of theory testing and leads to the development of

data-driven models that cannot be theoretically justified and would rarely replicate on

different samples. This is because modification indices capture purely *statistical* adjustments

to a model based on the idiosyncrasies of the *specific* sample at hand and do not consider

whether such adjustments also make *theoretical* sense. For example, modification indices

may suggest introducing error covariances or allowing cross-loadings of indicators,

adjustments which in the vast majority of cases cannot be theoretically justified. As a result,

one of SEM's biggest strengths (i.e., identification of a model's key misspecifications) has

unfortunately become also one of its big weaknesses.

Bearing the above in mind, the weight is with the author to provide (and with the

reviewer to require) convincing theoretical justifications not only for the hypothesized paths

in a model but also for the non-hypothesized ones (i.e., zero paths). Moreover, authors should

not modify their models post-hoc unless they (1) explicitly underscore the exploratory nature

of their work, and/or (2) are able to replicate their revised model on fresh data from

subsequent studies and, in particular, test the previously identified paths by modification

indices as free parameters.

2.9. Question 9: What are nested models?

Imagine two SEM models (e.g., Model A and Model B) that have exactly the same manifest

and latent variables. Now, assume that in Model A the relationships among latent variables

are set free to be estimated while in Model B some relationships between latent variables have been constrained by setting them to zero. In this case, we would call B a *model nested* in A, because Model B can be obtained after constraining one or more parameters of Model A to a fix value (e.g., zero) or setting them equal to some other parameter. There can be several models nested within a particular model because there are numerous types and number of parameters that can be constrained, leading to several nested models that correspond to different representations of the relationships among the model constructs.

Figure 6 presents several nested models. Model A is a model where all latent variables are directly connected with any other latent variable in the model (i.e., all potential structural paths are free to estimate). Models B1, B2 and B3 are three models which are all nested in Model A. Model A reduces to Model B1 if the paths (b), (e), and (f) are constrained to zero. Thus, assuming that the same indicators have been used to measure the latent variables in both models, Model B1 has three degrees of freedom more than Model A. Model A also reduces to Model B2 if path (c) is constrained to zero (resulting in one degree of freedom more than Model A). Finally, Model A reduces to Model B3 by constraining paths (b) and (e) to zero.[11] Note that Models B1-B3 represent different theoretical propositions. Model B1 is a serial mediation model where the impact supplier's staff warmth on repurchase intent is expected to be fully mediated through supplier's staff competence and satisfaction with the supplier. Model B2 suggest that only supplier staff warmth (but not competence) impacts satisfaction with the supplier. And Model B3 suggests that satisfaction with the supplier fully mediates the impact of both supplier staff competence and warmth on repurchase intentions.

Insert Figure 6 about here

---

[11] Model B3 is very similar to the model in Figure 2, the only difference is that supplier staff warmth is now modeled as an antecedent of supplier staff competence.

To sum up, given two models, as long as (1) the same variables are included, and (2) one model can be obtained through restricting one or more parameters of the other, the models are nested and thus formal fit comparisons can show which of the two receives stronger empirical support with the data at hand. Such fit comparisons employ what are known as *chi-square difference ($\Delta\chi^2$) tests* and involve subtracting the χ^2 value of the less restricted model from the χ^2 value of the more restricted model. This difference is also distributed as a χ^2 statistic with degrees of freedom equal to the difference in the degrees of freedom of the two nested models.

Note that nested model comparisons often provide ammunition to authors when counteracting reviewer comments that cast doubt on their theoretical model setup. Quite often reviewers doubt the need for the inclusion of some structural paths or come up with alternative theoretical models to the authors' proposed ones. If these *rival models* are nested within the same overall model as the authors' originally proposed model, authors can engage in formal chi-square comparisons and establish on empirical grounds which of the proposed rival models is more consistent with the data at hand.

2.10. Question 10: What are equivalent models?

Two models are called *equivalent* when they include the same observed variables but also have the same number of constrained and free parameters (and thus the same number of degrees of freedom). In other words, equivalent models differ only in terms of their *model structures* (i.e., in terms of *which* specific parameters are set free or are constrained). Despite having different theoretical setups and implications, equivalent models have *identical* overall model fit (and thus cannot be compared by means of chi-square difference ($\Delta\chi^2$) tests as is the case for nested models – see Question 9). Thus, choice among equivalent models must be made based on their theoretical plausibility rather than statistical criteria.

An example of two equivalent models is shown in Figure 7. One researcher could theoretically argue that the correct model is Model A where supplier staff warmth acts as a causal antecedent of supplier staff competence which, in turn, influences the level of satisfaction with the supplier and, through it, intention to repurchase. Another researcher could instead argue that it is competence what precedes warmth in the causal chain toward satisfaction and repurchase intent and, therefore, that the "correct" model is Model B. Testing which of these two mediation models is superior is not possible using chi-square comparisons for the simple reason that these two models have the same model fit and the same degrees of freedom.

Insert Figure 7 about here

The existence of equivalent models should serve as a reminder to researchers that their models are *not* unique in terms of their fit to empirical data and that there can be *many* other models with the same variables but different structures that will produce exactly the same fit. Thus, just because a well-fitting model is obtained authors should not automatically assume that theirs is the *only* model or the "true" model that is consistent with the empirical data at hand.

3. Conclusion

SEM is a powerful analytical technique that has been increasingly attractive to IMM researchers over the last years. Nevertheless, IMM authors often struggle in understanding, using, and reporting the results of their SEM analyses appropriately. This results in intentional or unintentional misuse of the method and, subsequently, threatens the validity of published research findings. The purpose of the present paper was to provide a common

denominator for every IMM researcher using SEM by offering some non-technical explanations to key concepts of the method and thus effectively provide the common ground needed for the IMM community to safeguard its research contributions.

In summarising the main takeaways from the issues discussed above, we develop a list of guidelines for authors and reviewers/editors dealing with SEM-based manuscripts to achieve the highest possible accuracy and transparency when reporting or assessing the results of SEM applications (Table 2). These guidelines directly correspond to the ten fundamental concepts of SEM discussed in this article. Although this list is by no means exhaustive and it is expected that both authors and reviewers will delve into higher levels of analytical detail in their exchanges, we believe these guidelines represent the minimum level of reporting that a SEM-based manuscript should exhibit for an adequate representation of authors' efforts and a fair account of reviewers' requests. Although most manuscripts are expected to satisfy at least some of the guidelines presented in Table 2, we suggest that its contents are used as a checklist by authors before final manuscript submission and by reviewers as a reminder of areas where mistakes or inappropriate reporting are likely to take place.

Insert Table 2 about here

References

Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335-358.

Diamantopoulos, A., & Riefler, P. (2011). Using formative measures in international marketing models: A cautionary tale using consumer animosity as an example. *Advances in International Marketing*, *10*(22), 11-30.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.

Holbert, R. L., & Stephenson, M. T. (2002). Structural equation modeling in the communication sciences, 1995–2000. *Human Communication Research*, *28*(4), 531-551.

Hult, G. T. M., Ketchen, D. J., Shaojie Cui, A., Prud'homme, A. M., Seggie, S. H., Stanko, M. A., Xu, A. S., & Tamer Cavusgil, S. (2006). An assessment of the use of structural equation modeling in international business research. In *Research methodology in strategy and management* (pp. 385-415). Emerald Group Publishing Limited.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90-98.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kumar, V., Sharma, A., & Gupta, S. (2017). Accessing the influence of strategic marketing research on generating impact: Moderating roles of models, journals, and estimation approaches. *Journal of the Academy of Marketing Science*, *45*(2), 164-185.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226.

Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *46*(6), 1148-1172.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. *Handbook of market research*, *26*, 1-40.

Schumacker, R. E., & Lomax, R. G. (2016). *Review of a beginner's guide to structural equation modeling*. New York: Routledge.

Steenkamp, J. B. E., & Van Trijp, H. C. (1991). The use of LISREL in validating marketing constructs. *International Journal of Research in Marketing*, *8*(4), 283-299.
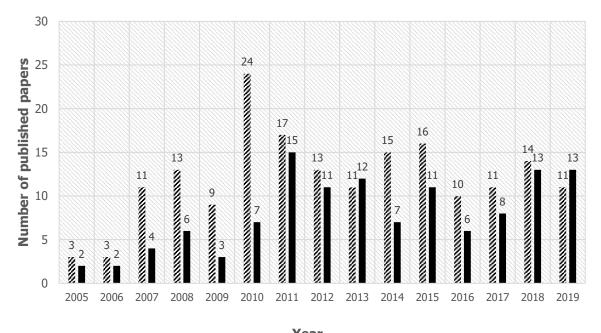
Figure 1: Published papers using Structural Equations Modeling in *Industrial Marketing Management* (2005-2019)

Figure 2: An illustrative structural model

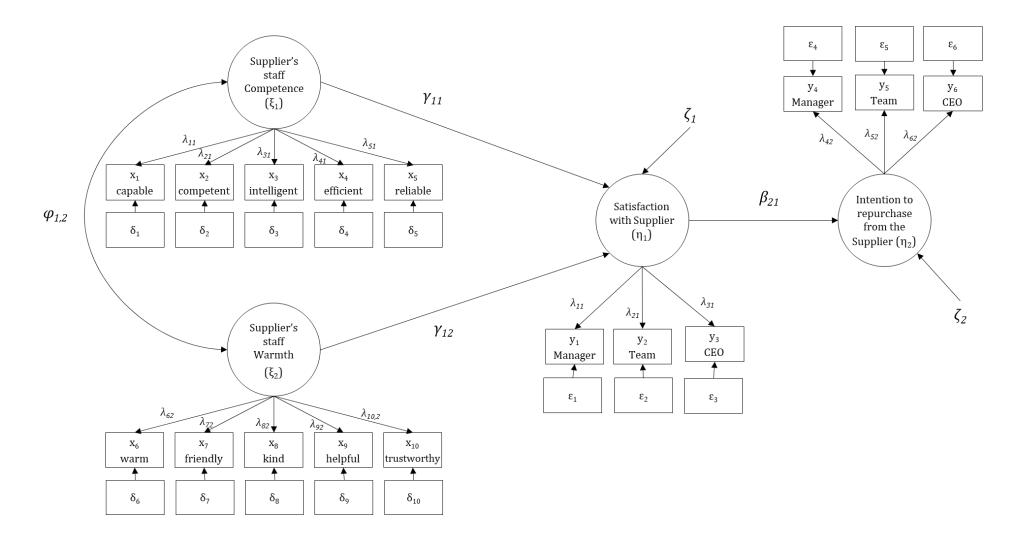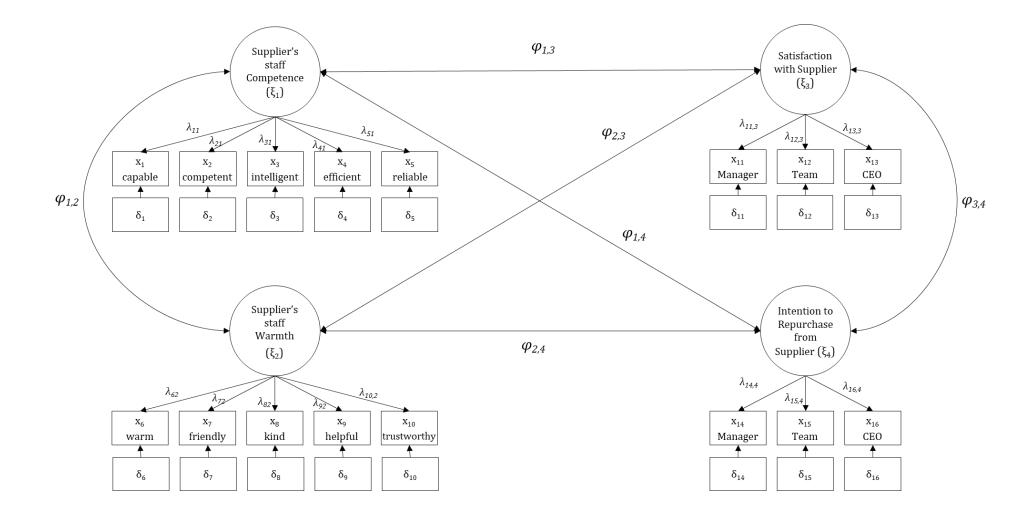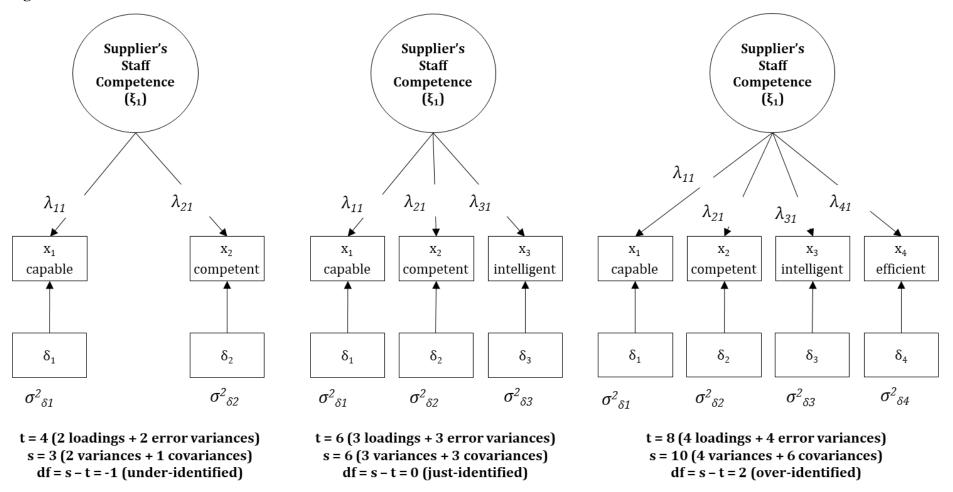# Figure 3: An illustrative measurement model

Figure 4: Model identification

$t = 4$ (2 loadings + 2 error variances)
$s = 3$ (2 variances + 1 covariances)
$df = s - t = -1$ (under-identified)

$t = 6$ (3 loadings + 3 error variances)
$s = 6$ (3 variances + 3 covariances)
$df = s - t = 0$ (just-identified)

$t = 8$ (4 loadings + 4 error variances)
$s = 10$ (4 variances + 6 covariances)
$df = s - t = 2$ (over-identified)

Note: Assumes the latent variable variance (φ_{11}) has been fixed to 1 for scaling purposes.
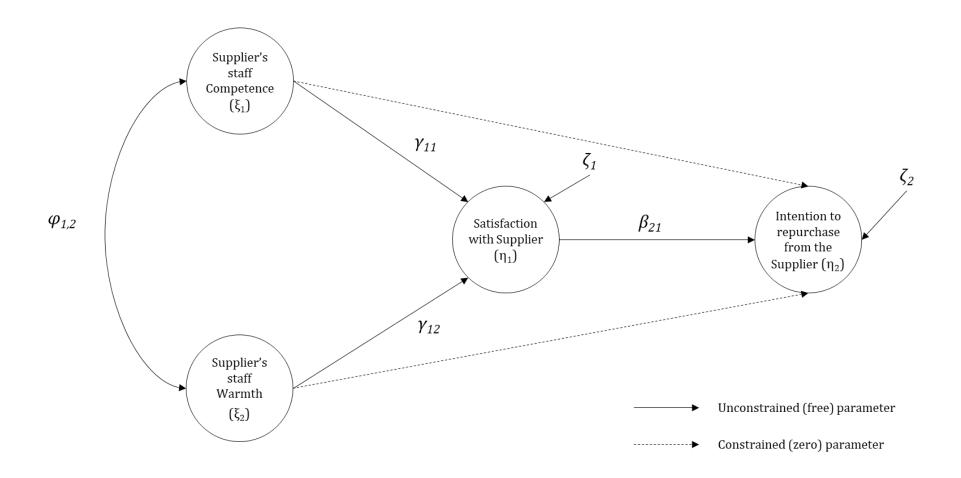
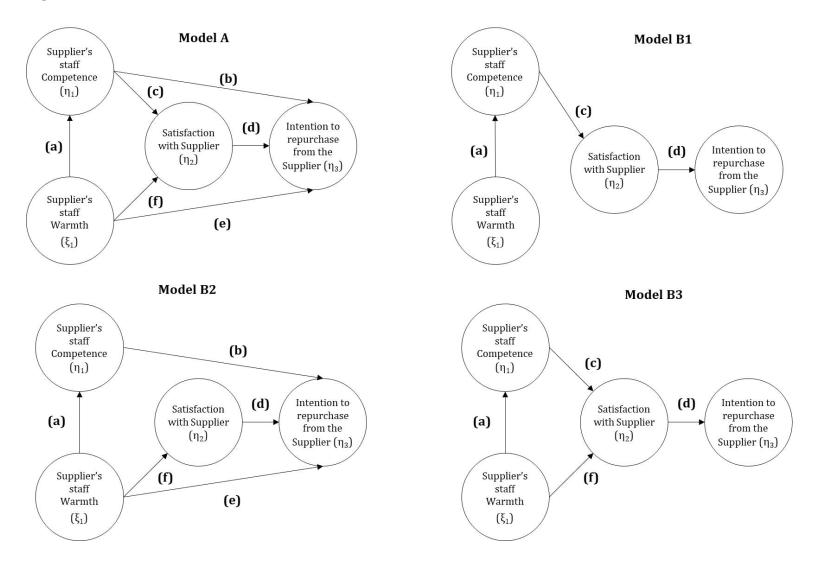**Figure 5: Constrained and unconstrained (free) parameters**
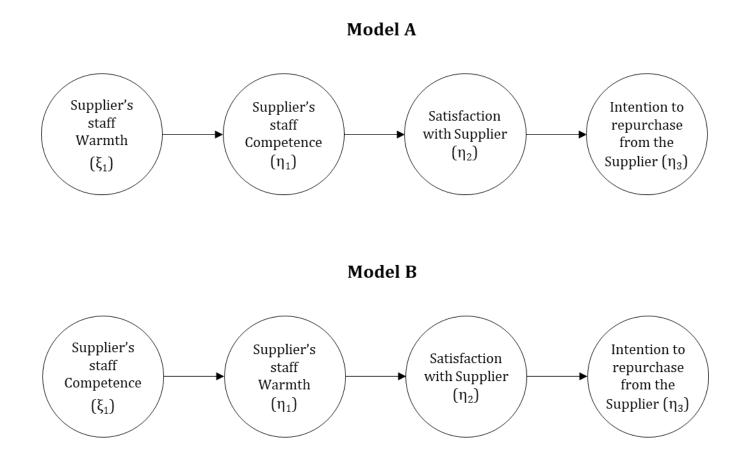
Figure 6: Nested models

Figure 7: Equivalent models

Model A

Model B

Table 1: SEM terms and notation

Term	Definition - description	Notation
Exogenous latent variable	A latent (unobservable) variable that is used to predict other variable(s) in the model; exogenous variables are not explained by any other variable in the model (no arrows leading to endogenous variables).	ξ (KSI)
Endogenous latent variable	A latent (unobservable) variable that is being predicted by the exogenous variable(s). There is at least one exogenous variable predicting it. It can also be predictor of other endogenous variables (in which case it is also called a mediator).	η (ETA)
Manifest variable (indicator)	A measured (observed) variable used as an indicator to measure some latent (unobservable) variable in the model. It can be an indicator of either endogenous or exogenous latent variables.	x (exogenous variable indicator) y (endogenous variable indicator)
Endogenous variable error	The error (i.e., random disturbances) in the endogenous variable capturing variance unexplained by the variable's predictors.	ζ (ZETA; error of η)
Manifest variable error	The measurement error of the indicators (items) used to measure endogenous or exogenous latent variables.	δ (DELTA; error of x) ε (EPSILON; error of y)
Indicator loading	The regressions weight linking manifest indicators to latent constructs (for reflectively-measured constructs). It shows how well the latent construct explains the indicator's variance. It helps assessing the measurement model, the appropriateness of individual indicators and the psychometric properties of latent variables.	λ (LAMDA)

Causal relationship	A theoretical, directiuonal relationship between any two latent variables in the model. It can be a relationship between one exogenous and one endogenous variable or between two endogenous variables. It is depicted by a line with an arrow end entering the endogenous variable (unidirectional relationship).	γ (GAMMA) Exogenous variable → Endogenous variable β (BETA) Endogenous variable → Endogenous variable
Non-directional relationship	A correlational (non-causal) relationship between two latent exogenous variables in the model. It is depicted by a double-arrow linking two exogenous variables. It captures the covariance between exogenous variables.	φ (PHI)

Table 2: Guidelines to IMM authors and reviewers/editors

Issue	Advice to authors	Advice to reviewers / editors
Latent variables and scaling	▪ Provide a detailed description of all latent variables and their corresponding manifest indicators. ▪ Discuss your scaling procedure. ▪ Identify which manifest indicator is used as scaling indicator.	▪ Request a table with all manifest indicators and their psychometric properties (e.g., loadings, indicator reliabilities, etc.). ▪ Check whether the scaling indicator has face validity in light of the construct's conceptual definition and described content domain.
Parameter estimation	▪ Provide a full list of important (if not all) estimated parameters, accompanied by standard errors, t-values and significance levels. ▪ Report parameter estimates for both measurement and structural models.	▪ Assess the statistical and practical significance of reported parameters. ▪ Assess the validity of the measurement model before proceeding with the assessment of structural parameters. ▪ Ask for re-estimation of structural model if severe issues with the measurement model exist.
Standardized parameters	▪ Be consistent in your reporting of parameter estimates. ▪ If both standardized and non-standardized parameters are reported, be clear on what is what. ▪ Interpret the results appropriately depending on whether your report standardized or unstandardized parameters. ▪ Report effect sizes.	▪ Ask authors to clarify whether they report standardized or unstandardized parameters in text and/or in tables. ▪ Require effect size estimates. ▪ Consider the unit of analysis of the used scales and guide authors in reporting accordingly.
Constraints	▪ Design your model and identify how many parameters are free and how many parameters are constrained before testing. ▪ Make sure that you have a theoretical explanation for your constrained parameters.	▪ Ask authors to provide theoretical arguments for their decision to constraint a parameter that should not intuitively be constrained. ▪ Check for dubious fixed or free parameters in the model (e.g., free estimation of error covariances between manifest variables of different latent constructs, zero covariances between exogenous variables).
Model identification	▪ Calculate your degrees of freedom before collecting data or testing the model. ▪ In case of under-identified models, consider remedies such as restricting parameters that should theoretically be unrelated or search for latent construct scales with more items. ▪ Report degrees of freedom in the model results.	▪ Require the exact number of degrees of freedom for every estimated model. ▪ Check whether authors' reported degrees of freedom are in line with the described model setup and, if not, ask for a list of constrained parameters to identify discrepancies. ▪ Assess the sample size in light of these degrees of freedom.

Independence / Saturated models	• Check and potentially report model fit statistics of the independence and saturated models.	• Ask for model fit statistics of the independence and saturated models if needed.
Model fit	• Include both global and local fit statistics. • When discussing the appropriateness of the overall model do not rely only on one fit index.	• Require discussion of more than one fit statistics. • Check whether the reported indices are above the required thresholds. • Be aware of the sensitivities of each fit index (e.g., sample size, degrees of freedom) and comment appropriately.
Model modification	• Do not include additional paths after the first estimation of the model unless a clear theoretical argument for their inclusion can be provided. • Be transparent on which paths or parameters were set free post-estimation. • When extensive modifications are proposed, consider testing the updated model on a new sample.	• Ask authors whether any modifications were made after the original model testing. • Ask for theoretical arguments for modifications and added paths across manuscript revisions. • Do not ask authors to consider inclusion of theoretically indefensible paths without strong reasons.
Nested models	• Test rival nested models if alternative theoretical possibilities exist or if asked by reviewers. • Use nested model comparisons to assess reasonable modifications such as inclusion of direct (on top of mediating) paths or to assess effect size differences between structural paths.	• Make a theoretical case for the nested/rival model asked from the authors. • Make sure that the any models presented as nested are indeed nested. • Do not ask authors to provide empirical comparisons between non-nested models using chi-square tests.
Equivalent models	• Do not treat equivalent models as nested. • Do not attempt chi-square comparisons between equivalent models as they are impossible. • Counter rival equivalent models using theoretical arguments.	• Do not ask authors to test among equivalent models in an empirical manner (e.g., chi-square comparisons)