**A Framework for Using Structural Equations Modeling in Information Systems Research**

*Navid Aghakhani*

*The University of Tennessee at Chattanooga*

*navid-aghakhani@utc.edu*

*Beni Asllani*

*The University of Tennessee at Chattanooga*

*Beni-Asllani@utc.edu*

*Richard S. Allen*

*The University of Tennessee at Chattanooga*

*Rich-Allen@utc.edu*

*Contact Author: Rich Allen*

***Abstract*** – *The use of Structural Equation Modeling (SEM) in behavioral science and particularly in Information Systems (IS) research is growing rapidly. By using SEM, researchers can theoretically represent and empirically test models that contain latent variables. Despite the great interest in the use of SEM in IS research, sufficient guidelines have not been established for addressing the criteria for selecting either Covariance Based SEM (CB-SEM) or Partial Least Square SEM (PLS-SEM). Our study fills this gap by proposing a framework to use when selecting the appropriate form of SEM to use. We also highlight the various psychometrics properties and the statistical tests required to be reported in various stages of SEM Analysis. Our framework is then applied to a sample of top tier IS publications which utilized SEM. Findings indicate that many of these studies did not clearly justify and explain their type of research as exploratory or confirmatory in nature, nor did they use the proper SEM method. Moreover, most of the articles failed to accurately report the psychometric properties of their measurement model and the required statistical tests. The framework proposed in this paper has the potential to significantly enhance the rigor and quality of future IS research which utilizes SEM.*

Keywords: Structural Equation Modeling, Partial Least Squares, Covariance Based, Research Methods, Information Systems Research

**A Framework for Using Structural Equations Modeling in Information Systems Research**

**Introduction**

The Information Systems (IS) discipline has witnessed a tremendous increase in the use of Structural Equations Modeling (SEM) as a method of data analysis (Rouse and Corbitt 2008). In principle, SEM has many advantages over General Linear Models (GLM). First, by incorporating path analysis, SEM allows more flexibility in exemplifying the relationships among theoretical constructs and enables endogenous variables to serve both as dependent and independent variables. Second, like factor analysis, SEM allows researchers to accommodate conceptual constructs that are recognized to be the primary cause of observed variables. Finally, SEM combines measurement and path modeling into a simultaneous assessment, which allows better estimates than GLM (Gefen et al. 2011; Golob 2003). The prototypical structural equation model can be conceptualized as a marriage of path analysis with factor analysis (Golob 2003).

There are two basic types of SEM; the first is Covariance Based SEM(CB-SEM) and the other is component-based Partial Least Square SEM (PLS-SEM). While SEM has been adopted as a widely used research method in social science, business disciplines are biased to use one of its variations. For instance, PLS-SEM has been widely used in the Information Sciences (IS) discipline (Sharma et al. 2019), however, the term SEM for many business disciplines is equivalent to CB-SEM (Ali et al. 2018). We argue that the appropriate SEM method should be used based on the type of study and not the prevalence of using a specific variation of SEM in a particular discipline. We believe the bias toward using PLS-SEM stems from a lack of coherent guidelines and criteria for choosing one method over the other based on which version is more appropriate to the specific research being conducted.

To address this gap, we propose an integrated framework for selecting the appropriate type of SEM by considering the type of study and the commonalities and differences between CB-SEM and PLS-SEM. Our integrated framework also clarifies the psychometric properties and the statistical tests required to be reported in SEM analysis based of the type of SEM being employed. To build the proposed framework, we rely on the research method papers published in top tier journals, specifically Management Information Systems Quarterly. To validate our framework, we analyzed a sample of 20 articles published in top-tier IS journals. One set of articles have an IT and firm performance theme which used CB-SEM, and another set have a user adoption and acceptance of technology focus which used PLS-SEM. Since we have selected papers that use SEM analysis, we consider the population of our sample to be relatively homogenous. As such, a sample size of about 12 papers is recommended as being sufficient to achieve a theoretical saturation (Boddy, 2016).

**Proposed Framework**

Structural equation modeling (SEM) has become a standard method of data analysis used in social science research disciplines including business research**.** SEM is used to analyze the structural relationships between measured variables and latent variables. Despite the prevalence of using SEM in business disciplines, there has been a lack of research methodology articles to help clarify the conditions under which the appropriate SEM variant should be used. This is very important, as some business disciplines tend to use one version of SEM more frequently than the other type. For instance, PLS-SEM is widely used in Information Systems and CB-SEM is mostly used in Marketing or Management disciplines. The aim of our proposed framework (see Figure 1) is to set some guidelines for business and more specifically IS researchers to use when deciding which form of SEM to use in their research. The proposed framework has different levels based on the various stages of research and data analysis. It also provides detailed information about psychometric and statistical tests required to be reported based on the type of SEM being employed.

**Figure 1: Framework of Using SEM in IS Studies**

****

**Level 1: Model Conceptualization**

The initial level of our framework is model conceptualization and refinement. This is primarily because the researchers need to have in-depth knowledge about the type of study they will be conducting and its theoretical underpinning before they can proceed with their research. This level has six steps:

**1.1 Determining the Type of Study**

PLS-SEM and CB-SEM are very different in their underlying philosophy, distributional assumptions, and estimation objectives. Ringle et al. (2012) highlighted the following reasons to choose PLS-SEM based on their survey:

1. Small Sample Size
2. Non-Normal Data
3. Formative Measures
4. Focus on Prediction
5. Model Complexity
6. Exploratory Research
7. Theory Development
8. Use of Categorical Variables
9. Convergence ensured.

Gefen et al. (2011) confirmed the above guidelines by stressing that with PLS-SEM the emphasis would be not on latent variables abstracted from reality but on prediction. This means a critical reason for using PLS-SEM should be based on predicting new relationships, theory development or an exploratory type of study. In other words, relying merely on incorporating formative indicators and non-normal data are not the sole reasons for using PLS-SEM over CB-SEM. For example, if the data depart from multivariate normality, we can still use CB-SEM (Quintana and Maxwell 1999) by employing some transformations (e.g. Logarithmic, square root, etc.); but, the researchers need to be careful in interpreting the results.

However, CB-SEM cannot accommodate formative indicators. Relying only on this issue is not a strong justification for using PLS-SEM over CB-SEM (MacKenzie et al. 2011). Therefore, we urge researchers to identify the type of study up front. In confirmatory studies, where the researcher wants to test a model that is grounded in strong theory, but is being investigated in new settings, they should select CB-SEM. In exploratory studies, where the purpose is to predict new relationships or develop new theories, the PLS form of SEM should be used (Hair et al. 2011).

**1.2 Construct Conceptualization**

MacKenzie et al. (2011) defined four important aspects of construct conceptualization. These include:

1. Examining how the focal construct has been used in prior research
2. Specifying the nature of the construct’s conceptual domain such as defining the entity (e.g. person, organization and so on) and property (e.g. feeling, perception and so on)
3. Specifying the conceptual theme of the construct (common or unique attributes, dimensionality and stability)
4. Defining the construct in unambiguous terms.

This important step forms the backbone of subsequent measurement development and model identification steps. We urge researchers to review the relative literature to come up with a clear concept of their main constructs as this topic is beyond the scope of this paper.

**1.3 Develop Measurements**

Once the construct is conceptualized, the next step is to produce the items (indicators) that denote the conceptual domain of a construct (MacKenzie et al. 2011). Researchers need to pay special attention to formative and reflective scales since misspecification in the measurement model may lead to misspecification in the structural model as well (Petter et al. 2007). Generally, reflective and formative construct scales are quite different.

For reflective constructs:

1. Indicators are manifestations of the construct
2. Changes in the indicator should not cause changes in the construct
3. Changes in the construct do cause changes in the indicators
4. Dropping an indicator should not alter the conceptual domain of the construct
5. Indicators are viewed as affected by the same underlying construct and are parallel measures that covary
6. Indicators are required to have the same antecedents and consequences and to have a high internal consistency and reliability

For formative constructs:

1. Indicators are defining characteristics of the construct
2. Changes in the indicators should cause changes in the construct
3. Changes in the construct do not cause changes in the indicator
4. Dropping an indicator may alter the conceptual domain of the construct
5. It is not necessary for indicators to covary with each other
6. Indicators are not required to have the same antecedents and consequences, nor have high internal consistency reliability (Cenfetelli and Bassellier 2009; Chin 1998b; Gefen et al. 2011; Jarvis et al. 2003; Petter et al. 2007)

**1.4 Collecting Data and Determining Sample Size**

The process of data collection is a broad topic and it demands an array of theoretical and methodological considerations, which are beyond the scope of this article. However, it is noteworthy to discuss sample size issues associated with using the two SEM techniques. In general, it is widely stated in the literature that PLS-SEM is more robust than CB-SEM when the sample size is small. However, it is important to note that even small sample sizes need to provide adequate statistical power (Ringle et al. 2012). In this sense, PLS-SEM is preferable to use in a situation where a model is complicated, and the sample size is rather small (Hair et al. 2011). In the case of CB-SEM, Kline (2015) suggests the N: q rule of thumb. N is the minimum sample size in terms of the ratio of cases and q is the number of parameters that need to be estimated. Although the minimum ratio of 20:1 is required in CB-SEM, in many studies, a sample size of at least 200 is found to be statistically acceptable (Quintana and Maxwell 1999).

**1.5 Specifying the Model**

According to Kline (2015), the representation of hypotheses in the form of structural equation modeling is an important specification. One aspect of this specification is related to the proper number of indicators per construct. In conducting a SEM study, there is a theoretical and statistical advantage to selecting multiple indicators. Methodologically, it is consistent with the post-positivistic philosophy of science and statistically, it is recommended to have three to four indicators per latent variable (Quintana and Maxwell 1999). Another aspect of the specification is related to the directionality of the hypothetical paths (Kline 2015; MacCallum and Austin 2000; Quintana and Maxwell 1999). Based on Kline (2015), directionality implies the following criteria:

1. Temporal Precedence
2. Association between covariance
3. Isolation (no other plausible explanation)
4. Correct effect priority (i.e., x causes y and not vice versa)

Given these criteria of causality, the nature of the path among endogenous and exogenous variables is subjected to theoretical consideration, measurement theory and methodological issues (Quintana and Maxwell, 1999). Consistent with these guidelines, the directional path in the experimental design can be specified. In addition, in longitudinal studies where variables are measured at different times, the directional path between variables measured at one time (e.g. pretest) and a second time (e.g. posttest) should be specified. The directional path is problematic in cross-sectional studies, yet, justifiable if the precedence of the causal variable is supported by existing theory and literature (MacCallum and Austin 2000; Quintana and Maxwell 1999). As stated earlier, identifying the constructs correctly and precisely is critical to conducting any SEM study. To this end, the researchers need to be aware of the omission of critical variables, as well as directionality as two potential sources of specification error in their study (Quintana and Maxwell 1999).

**1.6 Identifying the Model**

A model is identified when theoretically possible to drive the distinctive estimates for every parameter in the model (Kline 2015). Therefore, if a model is not identified, it should be re-specified (Kline 2015). There are some rules of thumb for model identification:

1. All recursive models are identified (Golob 2003; Kline 2015).
2. A model with zero degrees of freedom (saturated) is identified (Kline 2015).
3. It is possible to break a non-recursive model into blocks in which the relationships between blocks are recursive. When each separate block is identified then the entire model is identified (Golob 2003; Kline 2015).
4. If a standard Confirmatory Factor Analysis (CFA) model with a single factor has at least three indicators, the model is identified (Kline 2015).
5. If a standard CFA model with two or more factors has two indicators per factor, the model is identified (Kline 2015).

**Level 2: Analyzing the Measurement Model**

For the purpose of this study, we decided to divide this level into two parts. The first part emphasizes analyzing the measurement model of CB-SEM studies and the second part is devoted to PLS-SEM studies. The PLS-SEM and CB-SEM have many key differences regarding assumptions about the type and normality of the data, as well as the number of indicators and estimation methods being used. As stated in the previous section, the researchers need to identify the proper method of analysis prior to conducting any statistical analysis.

**2.1 Measurement Model Using CB-SEM**

There are various methods of generating parameter estimates such that the inconsistencies between estimated and observed covariance are minimized (Chin 1998a; Gefen et al. 2011). The most frequently used is the Maximum Likelihood (ML) estimation procedure (Quintana and Maxwell 1999). ML estimates have two major limitations. First, ML is based on a presumption of multivariate normality of the data. Second, ML only works well with medium to large sample sizes. In addition, it is worthwhile to bear in mind that one can incorporate reflective indicators with CB-SEM (Quintana and Maxwell 1999; Ringle et al. 2012). Analyzing the measurement model encompasses construct reliability issues as well as convergent and discriminant validity tests. It is important to note that we rely primarily on MacKenzie et al. (2011) to address these numeric measures.

**2.1.1 Construct Reliability**

According to Straub et al. (2004), construct reliability concerns the internal consistency of the measurement model and indicator reliability. There are essentially two tests to check for construct reliability. First, the internal consistency as measured by the Cronbach alpha of the items should be greater than 0.7. Second, Fornell and Larcker’s index of construct reliability should be greater than 0.7 (MacKenzie et al. 2011) and each individual loading should be greater than 0.7 (Hair et al. 2011).

**2.1.2 Convergent Validity**

Convergent validity will be achieved when items thought to reflect a construct show high correlations with one another when compared to the convergence of items relevant to other constructs regardless of SEM method being employed (Kline 2015; Straub et al. 2004). According to MacKenzie et al. (2011) and Hair et al. (2011), convergent validity will be satisfied if Average Variance Extracted (AVE) of the set of indicators is greater than 0.5.

**2.1.3 Discriminant Validity**

Discriminant validity tests if the concepts (factors) theorized to be distinct are in fact distinct. Kline (2015) recommends that if the correlation between two constructs is less than 0.9 they are considered distinct, but a more precise test is to check if the square root of the AVE for each latent variable is greater than its correction with other constructs. In addition, the loadings of the group of items on their conceptual construct must be greater than their cross-loadings on other constructs (Chin 1998a; Hair et al. 2011)

**2.1.4 Goodness of Fit**

As mentioned earlier, if a model with zero degrees of freedom is identified then it perfectly fits the data. An over-identified model will be more restricted (i.e. contain fewer paths). The degrees of freedom of an over identified model is equal to the number of paths that have been omitted from an identified model (Kline 2015). In practice, the Chi-Square statistic is used to test whether the omitted paths are equal to zero. Consequently, we are looking for insignificant differences (P-values greater than 0.1) for the Chi-Square statistic (Quintana and Maxwell 1999). But relying solely on the Chi-Square statistic has some important limitations. First, severe deviation from multivariate normality may result in model rejection. Second, if using a large sample size, the Chi-Square statistic will always reject the model, but a small sample size lacks statistical power (Golob 2003; Kline 2015; MacKenzie et al. 2011; Quintana and Maxwell 1999). Consequently, the following alternative fit indices have been proposed in the literature.

* Relative/normed chi-square (x2/df): Can be between 2 to 3 (Quintana and Maxwell 1999).
* Root mean square error of approximation (RMSEA): should be less than 0.06 (MacKenzie et al. 2011).
* Goodness of Fit (GFI) and Adjusted Goodness of Fit (AGFI): should be values greater than 0.95.
* Root mean square residual (RMR) and standardized root mean square residual (SRMR): should be less than 0.08 (MacKenzie et al. 2011).
* Normed-fit index (NFI): should be above 0.95 (MacKenzie et al. 2011).
* Parsimony fit indices: Parsimony-adjusted Goodness-of-Fit Index (PGFI) and Normed Fit Index (NFI) should be over 0.9 (Quintana and Maxwell 1999).
* There are many inconsistencies across the literature regarding which fit indexes should be reported in research articles. For the purpose of this article, we follow the advice from (Gefen et al. 2011; Kline 2015; MacKenzie et al. 2011) to report at least the Chi-Square test and the degrees of freedom, the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR).
	1. **Measurement Model Using PLS-SEM**

As stated earlier, PLS-SEM does not make any assumptions regarding the multivariate normality of the data, and this is one of PLS-SEM’s main strengths. In fact, even with CB-SEM, it is possible to use the Weighted Least Estimator (WSL) model parameter estimation method, which is more robust in dealing with non-normal data. However, Maximum Likelihood Estimation (MLE) is the more frequently used parameter estimate used with CB-SEM (Quintana and Maxwell, 1999). Conversely, PLS-SEM uses the Bootstrap method to estimate model parameters. Bootstrap essentially uses a small sample of data and estimates the parameters by replacement (Ringle et al. 2012). Evaluation of the measurement model with PLS-SEM is related to the type of measurements being used. For reflective and formative indicators, we have different procedures in terms of reliability and validity. The following sections elaborate the process involved in properly associating reflective and formative measurements.

**2.2.1 Reflective Indicators**

With PLS-SEM, evaluating measurement models in terms of reliability and validity of reflective indicators are like CB-SEM with a slight difference (Hair et al. 2011; MacKenzie et al. 2011). Like CB-SEM, researchers need to report and use the same metrics for construct reliability, convergent validity and discriminant validity. However, the model fit indices should not be reported for PLS-SEM studies because they are not a reliable measure of model fit.

In addition, there is a small difference between the tests that are used for measuring internal consistency. In the CB-SEM studies, Cronbach's alpha is used to measure the internal consistency while with PLS-SEM, Dillion-Goldstein’s rho is a better test because the Cronbach alpha assumes the tau equivalence (or parallelity) of the manifest variables, (i.e. each manifest variable is assumed to be equally important in defining the latent variable). Dillion-Goldstein’s rho does not make this assumption as it is based on the results from the model (i.e. the loadings) rather than the correlations observed between the manifest variables in the dataset (Chin 1998b).

**2.2.2 Formative Indicators**

The measurement of formative indicators requires a different consideration. The prominent reason is related to their conceptual definitions. In essence, with reflective measurements, constructs explain the variances found in the reflective indicators. Indicators are highly correlated and dropping an indicator does not change the conceptual dimension of the constructs. Therefore, the term AVE will be used to capture the variance caused by the construct in reflective indicators (Cenfetelli and Bassellier 2009; Chin 1998b). In terms of formative constructs, the term AVE is not relevant since the indicators explain the variance in a construct. They are not necessarily correlated since they try to explain different facets of a construct. These facts imply that the internal consistency and convergent validity are not relevant to formative measurements (Cenfetelli and Bassellier 2009; Petter et al. 2007). What is important about reliability in formative constructs are the significance weights (coefficients greater than 0.1) and their contribution to explain variance in the formative constructs (Cenfetelli and Bassellier 2009). In addition, multicollinearity is problematic in formative constructs since each indicator is expected to measure different facets of the construct; therefore, high inter-correlation is not expected, and it causes bias in the path estimates. The test for multicollinearity of indicators is to check if VIF is within the range of 3.3 to 10 (Cenfetelli and Bassellier 2009).

**Level 3: Analyzing the Structure Model**

The two important aspects of path analysis are R-square and the significance of path coefficients. During this phase, the researcher is attempting to identify a significant path in the structure model to support the hypothesis. Moreover, the high value of R-square indicates the amount of variance in one endogenous variable that is caused by the other endogenous or exogenous variables (Straub et al. 2004). Chin (1998b) recommends that the standardized path magnitude should be at least 0.20 and ideally above 0.30 to be considered meaningful. Ringle et al. (2012) argue that most of the research using PLS-SEM only relies on an R-square to illustrate the ability of the model to explain and predict the endogenous latent variables. They stress the use of a pseudo-F-test (f2 effect size) with PLS-SEM, which allows a researcher to examine the independent variable’s incremental explanation of a dependent variable and the cross-validated redundancy measure Q2 that better allows for an assessment of the model’s predictive relevance.

**Level 4: Specifying Alternative Models**

One of the significant limitations of SEM is that there are likely to be several alternative models that cannot be empirically distinguished from a researcher’s theoretical model (Chin 1998a). The presence of many of these alternative models may signify that the study should not be conducted using SEM. Moreover, several plausible alternative models should be specified a priori that run counter to the researchers’ favored model (Quintana and Maxwell, 1999).

**Level 5: Reporting SEM Results**

The last level of our framework is related to reporting the results of the study. Gefen et al. (2011) provided a very concise guideline of what to report in SEM analysis. We have incorporated these guidelines into our overall framework, which encompasses the reasons underlying using the methods, choice of estimation methods, dropping any indicators, reporting any new re-estimate and accurately reporting model fit.

**Table 1: Summary of SEM Levels**

|  |  |
| --- | --- |
| Levels | **Outlines** |
| Level 1 | **Level 1: Model Conceptualization and Refinement** |
| **Step 1: Type of study**The decision about the type of research (confirmatory vs. exploratory);(1) PLS is better for Exploratory and theory building.(2) CB-SEM is better for theory testing. |
| **Step 2: Construct conceptualization**(1) Examine how the focal construct has been used in prior research.(2) Specify the nature of the construct’s conceptual domain (i.e. entity and property). (3) Specify the conceptual theme of the construct (common and unique attributes, dimensionality and stability).(4) Define the construct in unambiguous terms. |
| **Step 3: Develop measurements**(1) Consider differences between reflective and formative measurements.(2) Misspecification in the measurement model may lead to misspecification in the structural model. |
| **Step 4: Data collection and sample size**Data collection has its own methodology and considerations; (1) PLS-SEM works well with small sample sizes with regard to providing statistical power. (2) CB-SEM is used with ML estimate: N:q rule of thumb, with a minimum ratio of 20:1 is necessary.  |
| **Step 5: Model Specification**(1) The proper number of indicators per construct.(2) Directionality of structural path. (3) Specification error due to omitting critical constructs or directionality. |
| **Step 6: Model Identification (only done for CB-SEM)**(1) Theoretically possible to drive the distinctive estimates for every parameter.(2) Recursive vs. non-recursive model. (3) Just identified versus over identified. |
| Level 2 | **Level 2: Analyzing the Measurement Model** |
| (1) Reliability: Indicator reliability and internal consistency for reflective (if no internal consistency, test for formative).(2) Convergent validity (only with reflective). (3) Discriminant validity. (4) Model fit indices for CB-SEM. |
| Level 3 |  **Level 3: Analyzing the Structure Model** |
| (1) Type of estimate method (bootstrap for PLS-SEM or ML for CB-SEM).(2) Significance and magnitude of path coefficients. (3) Evaluation of R-square. |
| Level 4 |  **Level 4: Specify Alternative Models** |
| The presence of a large number of alternative models may signify that the study should not be conducted using SEM. Moreover, a number of plausible alternative models that run counter to the researchers’ favored model should be specified a priori. |
| Level 5 |  **Level 5: Reporting the SEM Study** |
| (1) Include a justification of choosing method.(2) Report model re-estimation and omitted indicators with new scales.(3) Explain the choice of estimate method. (4) Report proper fit indices. |

**Analyzing Empirical Research Based on the Proposed Framework**

The purpose of this section is to use our proposed framework to analyze and criticize the use of SEM in a random sample of twenty empirical studies published in top-tier IS journals. Our sample covers major IS research topics such as information technology adoption, the business value of information technology and information security. We have done this analysis in two parts. In the first part, we perform a qualitative evaluation of ten CB-SEM empirical studies and we summarized the results in Table 2. In the second part, we performed the same analysis on the ten PLS-SEM based papers. We summarized these results in Table 3. We focus primarily on just the SEM data analysis aspects in both parts.

**Part 1: CB-SEM empirical studies**

**1.** Tanriverdi (2006) addressed under which conditions cross-unit IT synergy improves the firm performance. The nature of this study is both exploratory and confirmatory. It is exploratory because it develops new measurements for IT relatedness and confirmatory in the sense that it is built on theoretically proven relationships between constructs.

In line with our framework, the author first conceptualized the dimensions of synergy. The use of PLS-SEM is more appropriate since it develops new measurements; however, the data analysis uses LISREL for both confirmatory and exploratory analysis. The sample size (236) is consistent with the minimum requirement of data analysis using the SEM method. In addition, following our framework, the Cronbach alpha was used to measure internal consistency and it was greater than 0.7. The significance of the standard loadings was checked to measure convergent validity; however, the author should have reported the loadings (> 0.7) or AVE (> 0.5). In terms of discriminant validity, Chi-Square (X2) was used instead of using AVE or cross-loadings assessment.

Concerning the goodness of fit-model, only three out of four required indices were reported. SRMR should also have been reported. In addition, the cut-off measures are dated and newer studies demand a higher level of goodness of fit. In terms of the path model, the hypotheses were examined by evaluating the significance of path coefficients.

**2.** Tanriverdi (2005) followed the same approach when examining the mediating effects of IT relatedness on firm's performance through corporate capabilities. The nature of this study is confirmatory in that it tried to examine whether the established relationships hold in new settings. Consistent with our framework, he conceptualized cross-unit knowledge management capability. This study raises the same concerns discussed with Tanriverdi (2006) regarding the validity study of measurement and reporting of SRMR in the fit model.

**3.** Banker et al. (2006) addressed the same research question of Tanriverdi (2005) in that it investigates whether the Plant Information System indirectly affects plant performance through manufacturing capabilities. One important concern regarding this study is incorporating the formative relationships between controls and dependent variables that by default signify the use of PLS-SEM. Despite this important issue, the CB-SEM method was used to analyze the data. They used the square of indicators to evaluate indicator reliability; however, they did not drop the items with measures of less than 0.5. In addition, they failed to report whether the Cronbach alpha was used to measure internal consistency. Like the Tanriverdi papers, this study did not use AVE to measure convergent validity; instead, they relied solely on their evaluation of the significance of the T-values of the indicators' loadings. Discriminant validity was examined using the Chi-Square (X2) of a separate structural model that was not consistent with our framework. Unlike the Tanriverdi articles, the authors reported the choice of model estimation by highlighting the robustness of the WSL approach in working with non-normal data. In terms of model validity, SRMR was not reported. Regarding the path model, the hypotheses were examined by evaluating the significance of path coefficients.

**4.** Segars and Grover (1998) used the explanatory factor analysis to investigate the dimensions of Information Systems planning success. They used CFA to evaluate the dimensions of Information Systems planning success. In terms of reliability, Cronbach’s alpha was used to measure internal consistency. It appears that some items have been removed due to a lack of indicator reliability, but exactly how these items were evaluated was not elaborated. They used AVE > 0.5 to check for convergent validity and a crosscheck AVE with factor correlations for discriminant validity that is in line with our general framework. In terms of model fit, they failed to report CFI and RMSEA.

**5.** Kearns and Lederer (2000) applied the dimensions from Segars and Grover (1998) in the context of alignment between a business plan and an IS plan to help create competitive advantage. This study followed our framework to measure reliability and validity. They used the square of item loadings to measure indicator reliability and Cronbach's alpha to measure internal consistency. AVE was checked both to evaluate convergent validity (AVE > 0.5) and discriminant validity (AVE > LVs correlation). In terms of the model fit, they failed to report SRMR. Regarding the path model, the hypotheses were examined by evaluating the significance of path coefficients.

6. Peukert et al. (2019) investigated the drivers of consumers’ adoption of immersive shopping environments. Their study is one of the few studies that explicitly argued the use of CB-SEM over PLS-SEM due to the confirmatory research objectives. For internal consistency of items, the authors properly reported dropping one of the items that failed to exceed the 0.7 thresholds. However, they reported 0.6 as the required threshold for the standard loadings, which is inconsistent with the proposed framework. Their analyses of convergent and discriminant validity and model fit were consistent with our framework.

7. Aurigemma et al. (2019) adopted the theoretical lens of protection motivation theory (PMT) to examine voluntary password manager application use with individual home end-users. Since the purpose of their study was to replicate the core and full PMT nomologies in the new context, the use of CB-SEM was appropriate. For internal consistency of items, they used AVE, which is not consistent with our proposed framework. The authors reported the standard loadings of the constructs; however, some of the items were below the recommended threshold of 0.7. The analyses of convergent and discriminant validity were consistent with our framework. In terms of the model fit, the authors reported all of the required fit indices.

8. Budner et al. (2017) examined the drivers of continued usage of Information Systems (IS). Like Aurigemma et al. (2019) study, the research objectives were confirmatory, which implies using CB-SEM over PLS-SEM. The authors reported reliability and validity measures according to our framework. However, in terms of fit indices, they failed to report SRMR

9. Wang (2019) examined how external information sources affect users’ perceptions of cyber security risk. The author used CB-SEM to test the measurements and hypothesized models. However, the author should have used PLS-SEM because of the exploratory nature of this study. Regardless of this issue, the author reported the reliability and validity measures appropriately. In terms of fit indices, the author failed to report RMSEA.

10. Hwang and Cha (2018) examine whether employee’s security-related stress and roll stress could influence their information security compliance intention. Their study was exploratory in nature, which implies the use of PLS-SEM over CM-SEM. The authors reported the reliability and validity measures of the constructs consistent with our framework. The authors also properly reported Chi-Square, RMSEA, and CFI fit indices. However, they failed to report SRMR like most of the reviewed articles in this section.

**Table 2: CB-SEM Empirical Studies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | Measurement Model | Path Model |
|  |  |  | Reliability | Validity | Fit Model |  |
| Papers | Type of Indicators | Internal Consistency | Indicator Reliability | ConvergentValidity | DiscriminantValidity |  |  |
| Only for Reflective | Reflective |  Formative | OnlyReflective | Only CB-SEM | BothMethods | Only PLS |
| Reflective | Formative | Fornell and Larker | Cronbach’s alpha Alpha>0.7 | DG rho>0.7 | Standard loadings >0.7 | Square of loadings>0.5 | Significance of weights | VIF between 3.3 to 10 | AVE >0.5 | Significance of loadingsStandard loadings >0.7 | AVE > correlation between LVs | Indicator cross-loadings | Chi-Square test (p>0.1) and theDegree of freedom | RMSEA <0.06 | CFI >0.95 | SRMR<0.08 | Examine the Significance of pathcoefficients | R-square of the endogenous variable | Pseudo F-test (f2 effect size) | Cross-validated redundancyMeasure Q2 |
| Tanriverdi(2006) | X |  |  | X |  |  |  |  |  |  | X |  |  | X | X | X |  | X |  |  |  |
| Tanriverdi(2005) | X |  |  | X |  |  |  |  |  |  | X |  |  | X | X | X |  | X |  |  |  |
| Banker et al. (2006) | X | X |  | X |  |  | X |  |  |  | X |  |  | X | X | X |  | X |  |  |  |
| Segars and Grover (1998) | X |  |  | X |  |  |  |  |  | X |  | X |  | X |  |  | X |  |  |  |  |
| Keans and Lederer (2000) | X |  |  |  |  |  | X |  |  | X |  | X |  | X | X | X |  |  |  |  |  |
| Peukert et al (2019) | X |  |  | X |  |  |  |  |  | X | X |  |  | X | X | X | X | X |  |  |  |
| Aurigemma et al. (2019) | X |  |  | X |  |  |  |  |  | X | X |  |  | X | X | X | X | X |  |  |  |
| Budner et al. (2017) | X | X |  | X |  | X |  |  |  | X |  | X |  | X | X | X |  | X |  |  |  |
| Wang (2019) | X |  |  | X |  | X |  |  |  | X |  | X | X | X |  | X |  | X |  |  |  |
| Hwang and Cha (2018) | X |  |  | X |  | X |  |  |  | X |  | X | X | X | X | X |  | X |  |  |  |

**Part 2: PLS-SEM empirical studies**

**1.** Karahanna et al. (2006) addressed the importance of individual beliefs about the capability of technology by extending the context of capabilities to four distinct and separate constructs. Consistent with our framework, they provided the conceptual definition of the context of capability construct. However, they failed to report the type of measurement (indicators); however, they did use the criteria to test reflective indicators in terms of reliability and validity. In terms of reliability, internal consistency was examined using Cronbach's alpha; however, Dillon-Goldstein's rho would be a better measurement for internal consistency. In addition, they did not clarify the measurements for indicator reliability. In terms of convergent validity, they checked for the loadings of constructs. For discriminant validity, they examined if AVE is greater than LVs correlations. In terms of path analysis, the R-Square of each endogenous variable and the significance of path were examined.

**2.** Venkatesh and Morris (2000) examined the role of gender in the adoption of new technology. The concept of constructs was clarified in their study, but the authors did not provide the details regarding tests used for examining their measurement model. They clearly elaborated on the methods of discriminant validity (AVE is greater than LVs correlations) but the method of testing for convergent validity and contrast reliability were not reported even in the appendixes. Similar to the other studies, the R-Square of each endogenous variable and the significance of hypothetical paths were used to analyze the structure model.

**3.** Al-Gahtani et al. (2007) addressed the role of culture in the acceptance of IT. The model had been empirically supported in other cultures; therefore, the aim of this study was to confirm if the model applies in new settings. In this sense, the type of this research is confirmatory and the use of PLS-SEM here is questionable. They did not provide any methodological justification to support their choice to use PLS-SEM over CB-SEM in this setting. In fact, based on our framework, CB-SEM would be more robust with this confirmatory type of research. In terms of the measurement model, they did not clarify the method of testing internal consistency. This fact is very important because the preferable test for internal consistency is Dillon-Goldstein's rho. The convergent validity was supported by examining the factor loadings. In terms of discriminant validity, AVE was greater than LVs inter-correlations. Similar to the other studies, the R-square and significance of path coefficients were examined in the structural model.

**4.** Chwelos et al. (2001) addressed the key success factors of adopting EDI in organizations. They incorporated both reflective and formative indicators in their research. Consistent with our framework, they highlighted the main reasons for choosing PLS-SEM by stressing the exploratory nature of their research and incorporating both refractive and formative indicators. In their measurement model, they used a separate approach for evaluating reflective and formative constructs. For reflective constructs, they used Fornell statistics instead of Cronbach's alpha to measure internal consistency. The convergent and discriminant validity were examined by testing the loadings of reflective indicators and whether the square of AVE is greater than LVs loadings. For formative measurements, they examined the significance of weights to test indicator reliability; however, the VIF test for multicollinearity was not examined. In line with our framework, examining internal consistency and convergent validity was not relevant to formative constructs. In terms of the structure model, they examined the R-square and significance of loadings. In addition, all path confidence estimates were greater than 0.2.

5. Zhou et al. (2018) adopted the consumer service life cycle (CSLS) framework to investigate the antecedents of consumers’ perceived transparency toward B2C ecommerce websites and how the perceived transparency, in turn, affects consumers’ intention to purchase from the website. They used PLS-SEM to test the conceptual model, which is consistent with our framework due to the exploratory nature of the study. However, they did not specify the reason behind using PLS-SEM over CB-SEM. They used both reflective and formative constructs in their theoretical model. In terms of reflective indicators, they reported the required reliability and validity measures. The authors also properly reported the VIF and the significance of the weights for the formative indicators. To test the structure model, they reported the significance of the path model, the effect size, and the R-square of the endogenous variable.

6. Niehaves and Plattfaut (2014) examined the effects of the wide range of antecedents of internet usage among the elderly by adopting the theoretical lens of Unified Theory of Acceptance and use of Technology (UTAUT) as well as the Model of Adoption of Technology in Households (MATH). The research objectives of this study were confirmatory, which indicates the use of CB-SEM. However, due to the lack of sample size, the authors have used PLS-SEM to test the theoretical model. They also employed the centroid weighting scheme to reduce the risk of overestimating the factor loadings. The report of reliability and consistency measures are in line with our framework. However, in the structure model analysis, the authors failed to report the effect size.

7. Shen et al. (2019) examined the drivers of social commerce engagement. The research objectives of their study were exploratory, which implies the use of PLS-SEM. The authors also have used formative constructs in their research model, which also implies the proper use of PLS-SEM over CB-SEM. Like Zhou et al. (2018), the authors reported the required reliability and validity measures of the reflective and formative constructs completely. However, the authors failed to report the effect size in the path model analysis.

8. Seethamraju et al. (2018) used the theoretical lens of Unified Theory of Acceptance and Use of Technology (UTAUT) to investigate the drivers of the adoption and use of mobile-based IT adoption. The authors explicitly elaborated on the predictive nature of their study, which implies the use of PLS-SEM to test the theoretical model. The authors also properly used the reflective indicators to measure the constructs and reported the reliability and validity measures. Also in accordance with our framework, the study reported the R-square of the endogenous variable, effect size (f2) and the Cross-validated redundancy Measure (Q2).

9. Koohikamali et al. (2015) examined the factors that predict the location based social network applications (LB-SNAs) users select to disclose their location. The authors correctly used PLS-SEM due to predictive research objectives of their research. The authors used reflective constructs in their study. The report of reliability and validity measures are in line with our framework. In the path model analysis, however, they failed to report the effect size (f2) and Cross-validated redundancy Measure (Q2).

10. Yang et al. (2016) examined the effect of social consensus in product reviews by examining the drivers of perceived risk and its indirect effects on the behavioral purchase intention. The purpose of this study was exploratory, which implies the use of PLS-SEM. The authors reported reliability and validity measures in accordance with our framework. However, in the analysis of the path model, the study failed to report the effect size (f2) and Cross-validated redundancy Measure (Q2).

**Table 3: PLS-SEM Empirical Studies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | Measurement Model | Path Model |
|  |  |  | Reliability | Validity | Fit Model |  |
| Papers |  Type of Indicators | InternalConsistency | Indicator Reliability | Convergentvalidity | DiscriminantValidity |
| Only for Reflective | Reflective |  Formative | OnlyReflective | Only CB-SEM | BothMethods | Only PLS |
| Reflective | Formative | Cronbach’s alpha Alpha>0.7 | Fornell and Larcker | DG rho>0.7 | Standard loadings >0.7 | Square of loadings>0.5 | Significance of weights | VIF between 3.3 to 10 | AVE >0.5 | Significance of loadingsStandard loadings >0.7 | AVE > correlation between LVs | Indicator cross-loadings | Chi-Square test (p>0.1) and theDegree of freedom | RMSEA <0.06 | CFI >0.95 | SRMR<0.08 | Examine the Significance of pathcoefficients | R-square of the endogenous variable | Pseudo F-test (f2 effect size) | Cross-validated redundancyMeasure Q2 |
| KarahannaEt al. (2006) | X |  | X |  |  |  |  |  |  |  | X | X |  |  |  |  |  | X | X |  |  |
| Venkateshand Morris(2000) | X |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  | X | X |  |  |
| Al-Gahtaniet al. (2007) | X |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Chweloset al. (2000) | X | X |  | X |  |  |  | X |  |  | X | X |  |  |  |  |  | X | X |  |  |
| Zhou et al. (2018) | X | X | X |  |  | X |  | X | X | X | X | X | X |  |  |  |  | X | X | X |  |
| Niehaves and Plattfaut1 (2014) | X |  | X |  |  | X |  |  |  | X | X | X | X |  |  |  |  | X | X |  |  |
| Shen et al. (2019) | X | X | X |  |  | X |  | X | X | X | X | X | X |  |  |  |  | X | X |  |  |
| Seethamraju et al. (2018) | X |  | X |  |  | X |  |  |  | X | X | X | X |  |  |  |  | X | X | X | X |
| Koohikamali et al. (2015) | X |  | X |  |  | X |  |  |  | X | X | X | X |  |  |  |  | X | X |  |  |
| Yang et al. (2016) | X |  | X |  |  | X |  |  |  | X | X | X | X |  |  |  |  | X | X |  |  |

**Conclusion**

The evaluation of top tier IS empirical papers using our framework reveals some common inconsistencies and limitations with IS research using SEM. The key overall findings are summarized below:

* Researchers need to first clarify the purpose of their study so that they can then choose the proper type of SEM analysis. Among this sample of top tier IS journal articles, only Chwelos et al. (2001), Peukert et al. (2019) and Seethamraju et al. (2017) referred to the confirmatory or exploratory nature of their studies as justification for using the proper SEM method.
* We believe that the boundary between confirmatory and exploratory research is not well defined. In this sense, more accurate research method papers are needed to clarify these boundaries for IS scholars. For instance, the Al-Gahtani et al. (2007) study was confirmatory in that they examined whether an established model can apply in new settings rather than predicting new relationships. CB-SEM would provide a more robust psychometric estimate for confirmatory research.
* Most studies failed to report the complete psychometric properties of their measurement model and their appropriate tests. IS researchers need to better clarify the exact tests and metrics used to evaluate the reliability and validity of their measurements. In terms of model fit, authors need to include the fit model indices listed in our framework as the minimum requirement of reporting. In addition, researchers need to be more precise in reporting the type of tests for both internal consistency and indicator reliability. Also, the PLS-SEM studies need to follow the Ringle et al. (2012) guidelines to report f2 effect size and Q2 in their study.

In summary, we believe that the framework presented here can be helpful to future IS researchers who plan to use SEM in their research. Editors and reviewers can also utilize our framework to evaluate the rigor of SEM research papers. Our framework provides guidance from the model conceptualization phase all the way through to the reporting stage. We have reviewed a sample of top-tier IS research publications utilizing this model and were able to find common improvements with regards to the use of SEM that could have been made to existing, highly regarded research.

**References**

Al-Gahtani, Said S, Geoffrey S Hubona, and Jijie Wang, 2007. Information technology (IT) in Saudi Arabia: Culture and the acceptance and use of IT. *Information & Management* 44 (8): 681-691.

Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., and Ryu, K, 2018. An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management*. 30(1): 514-538.

Aurigemma, Salvatore, Thomas Mattson, and Lori NK Leonard, 2019. Evaluating the Core and Full Protection Motivation Theory Nomologies for the Voluntary Adoption of Password Manager Applications. *AIS Transactions on Replication Research* 5(1): 3.

Banker, Rajiv D, Indranil R Bardhan, Hsihui Chang, and Shu Lin, 2006. Plant information systems, manufacturing capabilities, and plant performance. *MIS Quarterly* 30(2): 315-337.

Boddy, Clive Roland, 2016. Sample size for qualitative research. *Qualitative Market Research:*

*An International Journal* 19(4): 426-432.

Budner, Pascal, Mareike Fischer, Christoph Rosenkranz, Dirk Basten, and Lars Terlecki, 2017. Information System Continuance Intention in the Context of Network Effects and Freemium Business Models: A Replication Study of Cloud Services in Germany. *AIS Transactions on Replication Research* 3(1): 172-187.

Cenfetelli, Ronald T, and Geneviève Bassellier, 2009. Interpretation of formative measurement in information systems research. *MIS Quarterly* 33(4): 689-707.

Chin, Wynne W. 1998a, Commentary: issues and opinion on structural equation modeling. *MIS Quarterly* 22(1): vii-xvi.

Chin, Wynne W. 1998b, The partial least squares approach to structural equation modeling. *Modern Methods for Business Research* 295(2): 295-336.

Chwelos, Paul, Izak Benbasat, and Albert S Dexter, 2001. Empirical test of an EDI adoption model. *Information Systems Research* 12(3): 304-321.

Gefen, D, DW Straub, and EE Rigdon, 2011. An update and extension to SEM guidelines for admnistrative and social science research. *MIS Quarterly* 35(2): iii-xiv.

Golob, Thomas F, 2003. Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological* 37(1): 1-25.

Hair, Joe F, Christian M Ringle, and Marko Sarstedt, 2011. PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice* 19(2): 139-152.

Hwang, Inho, and Oona Cha, 2018. Examining technostress creators and role stress as potential threats to employees' information security compliance. *Computers in Human Behavior* 81: 282-293.

Jarvis, Cheryl Burke, Scott B MacKenzie, and Philip M Podsakoff, 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30(2): 199-218.

Karahanna, Elena, Ritu Agarwal, and Corey M Angst, 2006. Reconceptualizing compatibility beliefs in technology acceptance research. *MIS Quarterly* 30(4): 781-804.

Kearns, Grover S, and Albert L Lederer, 2000. The effect of strategic alignment on the use of IS-based resources for competitive advantage. *The Journal of Strategic Information Systems* 9(4): 265-293.

Kline, Rex B. 2015. *Principles and Practice of Structural Equation Modeling*. Forth Ed. (New York Guilford publications).

Koohikamali, Mehrdad, Natalie Gerhart, and Mohammadreza Mousavizadeh, 2015. Location disclosure on LB-SNAs: The role of incentives on sharing behavior. *Decision Support Systems* 71: 78-87.

MacCallum, Robert C, and James T Austin, 2000. Applications of structural equation modeling in psychological research. *Annual Review of Psychology* 51(1): 201-226.

MacKenzie, Scott B, Philip M Podsakoff, and Nathan P Podsakoff, 2011. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly* 35(2): 293-334.

Niehaves, Björn, and Ralf Plattfaut, 2014. Internet adoption by the elderly: Employing IS technology acceptance theories for understanding the age-related digital divide. *European Journal of Information Systems* 23(6): 708-726.

Petter, Stacie, Detmar W Straub, and Arun Rai, 2007. Specifying formative constructs in information systems research. *MIS Quarterly* 31(1): 623-656.

Peukert, Christian, Jella Pfeiffer, Martin Meißner, Thies Pfeiffer, and Christof Weinhardt, 2019. Shopping in virtual reality stores: The influence of immersion on system adoption. *Journal of Management Information Systems* 36(3): 755-788.

Quintana, Stephen M, and Scott E Maxwell, 1999. Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist* 27(4): 485-527.

Ringle, Christian M, Marko Sarstedt, and Detmar Straub, 2012. A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly* 36(1): iii-xiv.

Rouse, Anne, and Brian Corbitt, 2008. There's SEM and “SEM”: A critique of the use of PLS regression in information systems research. *ACIS 2008 Proceedings*: 81.

Seethamraju, Ravi, Krishna Sundar Diatha, and Shashank Garg, 2018. Intention to use a mobile-based information technology solution for tuberculosis treatment monitoring–applying a UTAUT model. *Information Systems Frontiers* 20(1): 163-181.

Segars, Albert H, and Varun Grover, 1998. Strategic information systems planning success: an investigation of the construct and its measurement. *MIS Quarterly* 22(2): 139-163.

Shen, Xiao-Liang, Yang-Jun Li, Yongqiang Sun, Zhenjiao Chen, and Feng Wang, 2019. Understanding the role of technology attractiveness in promoting social commerce engagement: Moderating effect of personal interest. *Information & Management* 56(2): 294-305.

Sharma, P., Sarstedt, M., Shmueli, G., Kim, K., and Thiele, K, 2019. PLS-based model selection: The role of alternative explanations in information systems research. *Journal of the Association for Information Systems*. Forthcoming.

Straub, Detmar, Marie-Claude Boudreau, and David Gefen, 2004. Validation guidelines for IS positivist research. *Communications of the Association for Information Systems* 13(1): 24.

Tanriverdi, Hüseyin, 2005. Information technology relatedness, knowledge management capability, and performance of multibusiness firms. *MIS Quarterly* 29(2): 311-334.

Tanriverdi, Hüseyin, 2006. Performance effects of information technology synergies in multibusiness firms. *MIS Quarterly* 30(1): 57-77.

Venkatesh, Viswanath, and Michael G Morris. 2000. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly* 24(1): 115-139.

Wang, Edward Shih-Tse, 2019. Effects of brand awareness and social norms on user-perceived cyber privacy risk. *International Journal of Electronic Commerce* 23(2): 272-293.

Yang, Jing, Rathindra Sarathy, and JinKyu Lee, 2016. The effect of product review balance and volume on online shoppers' risk perception and purchase intention. *Decision Support Systems* 89: 66-76.

Zhou, Liying, Weiquan Wang, Jingjun David Xu, Tao Liu, and Jibao Gu, 2018. Perceived information transparency in B2C e-commerce: An empirical investigation. *Information & Management* 55(7): 912-927.