FORMATIVE VS. REFLECTIVE MEASUREMENT MODEL: GUIDELINES FOR STRUCTURAL EQUATION MODELING RESEARCH

MOHD HAFIZ HANAFIAH*

Faculty of Hotel and Tourism Management, Universiti Teknologi MARA, Selangor, Malaysia
*Corresponding author: hafizhanafiah@uitm.edu.my

ABSTRACT. Various social sciences researchers have always debated the operationalisation of formative or a reflective measurement in Partial Least Squares Structural Equation Modeling (PLS-SEM). This paper aims to offer guidance on formative and reflective measurement model assessment in PLS-SEM. First, this paper explores and discuss the similarities and differences between the formative and reflective measurement model. Next, this paper reviews the practice of measurement model assessment for formative and reflective construct based on the latest methodological background. Finally, this paper proposes a set of guidelines in classifying the formative and reflective constructs and the steps in assessing the formative and reflective measurement model. This paper addresses the issue of measurement misspecification in PLS-SEM assessment by providing logical guidelines for researchers. This paper also helps to explain and suggest appropriate PLS-SEM assessment for researchers as they plan future research projects.

1. INTRODUCTION

Partial Least Squares Structural Equation Modeling (PLS-SEM) is a second-generation data analysis technique in the family of structural equation modelling ([1]; [2]). Different from the SEM covariance-based groups, PLS-SEM is a prediction-oriented approach to SEM, usually used for exploratory research and also appropriate for confirmatory research ([3]; [4]). Lauro and Vinzi ([5]) suggested that PLS-SEM is particularly useful for causal-predictive analysis in situations of high complexity and low theoretical information availability. Meanwhile, other researchers used...
the PLS-SEM approach because of its advantages over the covariance approach ([2]; [3]). The benefits of this soft-modelling approach include its ability to account the theoretical conditions, measurement conditions, distributional considerations, and practical considerations ([3]). Besides, PLS-SEM is also an exploratory statistical tool that is able to process primary or secondary data ([6]). Meanwhile, other researchers claimed that the PLS-SEM approach is suitable with prediction-oriented objective, abnormal data distribution and accommodates small sample sizes ([7]; [8]; [9]; [10]). Table 1.1 illustrates how the PLS-SEM approach is compared to the Covariance-based Structural Equation Modeling (CB-SEM) approach.

Table 1.1. Comparing Partial Least Square (PLS) to Covariance-based (CB) approaches of SEM

<table>
<thead>
<tr>
<th>Criterion</th>
<th>PLS-SEM</th>
<th>CB-SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Objective</td>
<td>Prediction oriented</td>
<td>Parameter oriented</td>
</tr>
<tr>
<td>Approach</td>
<td>Variance</td>
<td>Covariance</td>
</tr>
<tr>
<td>Assumption</td>
<td>Non-parametric</td>
<td>Parametric</td>
</tr>
<tr>
<td>Implication</td>
<td>Optimal for prediction</td>
<td>Optimal for parameter estimation</td>
</tr>
<tr>
<td>Model complexity</td>
<td>Large complexity</td>
<td>Small to moderate complexity</td>
</tr>
<tr>
<td>Sample size</td>
<td>Minimum of 30-100</td>
<td>Based on power analysis</td>
</tr>
<tr>
<td>Software</td>
<td>SmartPLS, WarpPLS, PLS-Graph</td>
<td>Amos, Lisrel, MPlus</td>
</tr>
</tbody>
</table>

As shown above, the approach of PLS-SEM is to explain the variance, similar to basic regression analysis and therefore, it is essential to note that PLS-SEM also provides Coefficient of determination ($R^2$) values besides indicating the significant relationships that lie among the construct which are able to denote on the performance of the model such as how far the model is performing. Among the advantages possessed by PLS-SEM if to be compared with the basic regression is its ability in handling various independent variable at one time even when it displays multicollinearity ([3]; [6]). Besides, some of the assumptions on regression are also shared by PLS-SEM, i.e. the ones which concern on the outliers as well as nonlinear data relationships. Lastly, the PLS-SEM’s characteristics which include the minimal demands when it comes to measurement scales, sample size and also residual distributions allow it to be utilised
in the circumstances either the relationships exist or not, and it can also be utilised in suggesting
the propositions for the later testing ([6]; [7]).

PLS-SEM involves a two-step approach which revolves around the estimation of the
measurement model right before an analysis is done for the structural model. It is also known as
an iterative algorithm which has the ability in separately solving out the blocks of the
measurement model and later estimates the path coefficients in the structural model. This paper
focuses on the differences and assessment of the formative and reflective measurement models.

2. MEASUREMENT MODEL

A measurement model is a component of the general model where latent constructs are
prescribed. Measurement models, as discussed in the psychological, sociological and
management literature - identify various instances where reflective and formative measures
differ. The most common distinction between reflective and formative measures has to do with
the relationship that is present when it comes to the construct and its measurement items ([11];
[12]; [13]; [14]; [15]). Commonly, the reflective modes act as the indicator of causality from
constructs to measurement items, and it is the other way around for formative modes. Figure 1
below exhibits the differences between formative and reflective construct.

![Diagram showing differences between formative and reflective construct](image)

*Figure 2.1. Differences between formative and reflective construct*

In term of formative construct, the latent variable is considered a consequence of its
respective indicators ([16]) and because the latent variable is defined by its indicators,
changing/replacing a formative indicator will alter the meaning of the latent variable ([17]). Alternatively, in reflective construct, indicators are deemed as the consequences of the latent variable to which they belong, which means items are manifested by the construct ([2]; [18]; [19]). The use of reflective indicators is interchangeable, and to a certain extent, it can even be removed. Another critical differentiation between the two models is whether the measurement items possess any correlation. In reference to the formative model, all measurement items are not necessary to appear having a high correlation, while the reflective model stipulates that there is a need for all measurement items to be highly correlated.

3. **CONSTRUCT CLASSIFICATION**

As the research data was collected prior to the model specification, the next step involved classifying the constructs as either formative or reflective. While some scholars argue that no construct is inherently reflective or formative, others suggest that a construct must be either reflective or formative based on its theoretical considerations ([14]; [15]; [20]). The rationale of these theoretical considerations is to develop items that measure the actual construct. The choice of measurement would affect the content, parsimony and criterion validity of the measurement model ([13]). Other researchers suggested that usage of incorrect measurement model will undermine the content validity of constructs, misrepresents the structural relationships between them, and ultimately lowers the usefulness of the research findings ([14]; [15]; [20]).

After reviewing the works of literature, this study found that three criteria: (i) the nature of the construct, (ii) the direction of causality between the indicators and the latent construct; and (iii) indicators characteristics that are used to measure the construct, is applicable in classifying the research constructs into formative or reflective measurement model.

3.1. **Nature of the construct.** Based on a reflective model, the latent construct is present (in an absolute sense) independently of the measures ([21]). This aspect is in line with many businesses and related methodological studies that use reflective measurement ([2]; [22]). Alternatively, for the formative model, the latent construct is dependent based on constructive, operational or
instrumental interpretation ([23], [24]). It is vital to highlight that due to the fact that formative indicators define the latent variable, they are not interchangeable. However, it was found that only there were only limited examples of formative models included in the literature of social science, specifically with regards to secondary data ([14]; [25]; [26]). They argued that secondary data tends to be very descriptive, may be challenging to obtain, and most of the time, it may not measure all the variables that are important to the research construct. Despite the limitations, they also supported that secondary data allows the researchers in testing complex hypotheses which involve multiple variables as well as large samples that act in facilitating the use of statistical techniques (e.g., structural equation modelling).

3.2. The direction of causality. The direction of causality between the construct and the indicators is the second consideration in deciding whether the measurement model is reflective or formative ([13]; [16]). Reflective models assume that the flow of causality flows is from construct to the indicators. Hence when there is a change in the construct, there will be a change in the indicators as well. Meanwhile, the reverse is true for formative models, where causality flows from the indicators to the particular construct. When there is a change in the indicators, it will result in a change in the construct under study. Also, it is essential to note that different causal direction can contribute towards significant implications in terms of the measurement error as well as the model estimation ([13]). Formative and reflective models were also found to have the main difference of which is basically on the treatment of measurement error, which then may affect the model estimation result.

3.3. Characteristics of indicators. Finally, in confirming whether the measurement model is reflective or formative, the differences with regards to specific indicator characteristics need to be analysed. For a reflective model, the content validity of the construct is not triggered by the inclusion or exclusion of one or even more indicators outside a domain. The indicators are interchangeable as they shared a common theme ([4]; [27]; [28]). However, in the case of formative models, types of indicators representing the construct as well as the number of constructs affect
the constructs itself, and thus, the conceptual meaning of the construct can change there is an addition or removal of an indicator. In this case, if the indicators represent the model conceptually, they are still considered adequate in the viewpoint of the empirical prediction.

Based on the above measurement and theoretical considerations, research constructs can be classified to either formative or reflective measurement model ([13]; [14]; [15]). Table 3.1 describes the justification process used to determine which constructs were reflective and which ones were formative.

Table 3.1. Formative and Reflective Construct Assessments

<table>
<thead>
<tr>
<th>The nature of the construct</th>
<th>Direction of causality</th>
<th>Characteristics of indicators</th>
<th>Verdict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent construct is dependent upon a constructive, operational or instrumental interpretation</td>
<td>Causality flows from the indicators to the construct; a change in the indicators results in a change in the construct understudy</td>
<td>Construct is sensitive to the number and types of indicators representing it</td>
<td>Formative</td>
</tr>
<tr>
<td>Latent construct exists (in an absolute sense) independently of the measures</td>
<td>Causality flows from the construct to the indicators; a change in the construct causes a change in the indicators</td>
<td>Construct is not sensitive; does not materially alter the content validity of the construct</td>
<td>Reflective</td>
</tr>
</tbody>
</table>

4. REFLECTIVE MEASUREMENT MODEL ASSESSMENT

Reflective measurement specifies that a latent or unobservable concept causes variation in a set of observable indicators, which therefore can be used to gain an indirect measurement of the concept. In order to examine the reflective measurement models, four parameters were examined: (i) internal consistency reliability, (ii) indicator reliability, (iii) convergent validity and (iv) discriminant validity ([3]; [9]; [29]). The criteria for the reflective measurement model fitting are presented below in Table 4.1.
Table 4.1. Reflective Outer Model Assessments

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Recommendations/Rules of thumb / Thresholds</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency reliability</td>
<td>Do not use Cronbach’s alpha; composite reliability &gt; 0.70</td>
<td>Bagozzi and Yi (1988) [18]</td>
</tr>
<tr>
<td>Indicator reliability</td>
<td>Standardized indicator loadings &gt; 0.70; in exploratory studies, loadings of 0.40 are acceptable</td>
<td>Hulland (1999) [28]</td>
</tr>
<tr>
<td>Convergent validity</td>
<td>Average variance extracted (AVE) &gt; 0.50</td>
<td>Bagozzi and Yi (1988) [18]</td>
</tr>
<tr>
<td>Discriminant validity -</td>
<td>Each construct’s AVE should be higher than its squared correlation with any other construct</td>
<td>Fornell and Larcker (1981) [30]</td>
</tr>
<tr>
<td>Fornell-Larcker criterion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross loadings</td>
<td>Each indicator should load highest on the construct it is intended to measure</td>
<td>Chin (1999) [7]</td>
</tr>
<tr>
<td>Heterotrait-monotrait ratio of</td>
<td>No discriminant validity problems (HTMT&gt;0.85 criterions)</td>
<td>Henseler et al. (2009) [8]</td>
</tr>
<tr>
<td>correlations (HTMT)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A threshold value of 0.70 was applied in assessing the internal consistency of the model specifically in the effort to determine the item’s minimum factor loadings (18). Measurements with loadings lesser than 0.70 were removed in cases where failure to eliminate them may contribute towards the increase of composite reliability that is greater than the threshold value ([3]). Meanwhile, the convergent validity was determined using the widely accepted method ‘average variance extracted (AVE)’ ([3]). The AVE value indicates that; on average, each construct can explain more than half of the variance of its measuring items and must be more than 0.50 ([14]; [18]).

Fornell and Larcker ([30]) criteria were used in examining the discriminant validity at the construct-level, whereas the discriminant validity at the item level was examined using Chin’s criteria ([7]). Implementing this two-fold technique in testing for the discriminant validity is supported by various researchers, as they suggested that the variance extracted estimates should be greater than the squared correlation estimate ([8]; [9]; [13]; [18]). Lately, many researchers proposed the Heterotrait-Monotrait ratio of correlations (HTMT) to assess discriminant validity in PLS-SEM ([31]). The HTMT can achieve higher specificity and sensitivity rates compared to the
cross-loadings and Fornell-Lacker criterion. From the HTMT results, if the HTMT values are less than 0.85, it indicated no discriminant validity problems and implied that the HTMT criterion did not detect the collinearity problems among the latent constructs ([31]).

5. FORMATIVE MEASUREMENT MODEL ASSESSMENT

Formative measurement specifies that the observable indicators are considered to cause the latent construct. Thus, formative constructs should be assessed based on the statistical significance and size of the indicator weights as well as collinearity among indicators ([6]). For the evaluation of the formative measurement model, this study adopted the guidelines outlined by ([6]; [20]; [32]). Specifically, three parameters should be examined: (i) multicollinearity; (ii) construct validity; and (iii) indicator reliability. The criteria for the formative measurement model fitting are presented below in Table 5.1.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Recommendations/Rules of thumb /Thresholds</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multicollinearity</td>
<td>Variance inflation factor (VIF) is used to determine whether there was high correlation between the formative indicators</td>
<td>Hair et al. (2017) [6]</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>Estimate the indicator weights to measure the contribution of each formative indicator to the variance of the latent variable.</td>
<td>Petter et al. (2007) [33]</td>
</tr>
<tr>
<td>Indicator Reliability</td>
<td>Calculates the outer loadings of the formative construct; if the item loadings are relatively high (&gt; .50), the indicator should be retained</td>
<td>Hair et al. (2012) [9]</td>
</tr>
</tbody>
</table>

Different to the reflective indicator by which moderate multicollinearity between construct indicators is desirable, the balance of the model can be triggered by the high multicollinearity present in the formative construct ([15]; [32]; [33]). Multicollinearity occurs when two or more predictors in the model are correlated and provide redundant information about the response. In determining whether there was a too high correlation between the formative indicators, many studies used the variance inflation factor (VIF). VIF is the reciprocal of the tolerance value; small VIF values indicate low correlation among variables and vice-versa.
The higher the value, the greater the correlation of the variable with other variables ([34]). For formative measures, there is a rule of thumb that clarifies; if VIF values are greater than 5, thus it represents high multicollinearity ([20]). Recently, other researchers recommended that multicollinearity exists if the VIF value is higher than 10 ([6]; [31]).

Next, it is important to note that the Cronbach’s alpha and the composite reliability will not be estimated, as much as formative indicators are not internally consistent ([7]; [35]). Moreover, the AVEs were not calculated, given the assumption that formative indicators demonstrate convergent validity ([35]). Therefore, in order to test for the construct validity, the estimation of the indicator weights in measuring the contribution of every each of the formative indicators to the variance of the latent variable should be applied. The item weights indicate whether or not an indicator can explain a significant portion of the variance of a formative construct ([36]; [37]). This step is in line with other researchers who suggested that indicator weights can be used to test the construct validity ([12]; [13]; [15]).

Lastly, researchers should also look at indicator reliability. The outer loadings of a formative construct should be tested to confirm the indicator reliability. When an indicator’s weight is not significant, but the corresponding item loadings are relatively high (> .70), the indicator should be retained, as been proposed by researchers ([9]; [29]). This will ensure that measurements are prioritised according to their reliability with regard to making estimations ([15]; [38]).

6. **PROPOSITION**

In any study, it is vital to acknowledge the different types of measurement models and understand the criteria involved when it comes to determining the measurement models’ mode. The formulation of the measurement model depends on the direction of the relationships that are present in reference to latent variables as well as the corresponding manifest variables. In general, there are distinctive types of measurement model that are available namely, (i) the reflective model or also called outward-directed model; and (ii) the formative model or also called the directed model. The general distinction between measurement model for reflective and formative measures must be distinguished. It is important to note that reflective measures generally represent the causality from the constructs to the specified measurement items, and meanwhile,
the formative measures consider the opposite. This study found that each measurement model must be tested by assessing the validity and reliability of the items and constructs used in each (reflective and formative) model. These specific steps must be taken to ensure that only reliable and valid constructs and measures are used, prior to assessing the nature of the relationships proposed by the research hypotheses. This study, therefore, suggests the following guidelines for researchers in assessing the differences between reflective and formative constructs as per Table 6.1 below.

**Table 6.1. Criteria used to Determine the Mode of Measurement Models**

<table>
<thead>
<tr>
<th>Reflective model</th>
<th>Formative model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using existing latent construct and involves a realist interpretation of a latent construct</td>
<td>Latent construct is formed, constructivist, operationalist, or instrumentalist interpretations</td>
</tr>
<tr>
<td>Causality from constructs to indicators; indicators are manifested by the construct; they are interchangeable and share a common theme; dropping an indicator does not alter the meaning of the construct</td>
<td>Causality from indicators to constructs; Indicators define the construct; they need not share a common theme; they are not interchangeable; dropping an indicator may alter the meaning of the construct</td>
</tr>
<tr>
<td>Measures have a high correlation, as they are all dependent on the same unobservable variable</td>
<td>Measures have positive, negative, low or zero correlation with one another</td>
</tr>
<tr>
<td>Measures have similar sign and significance of relationships with the antecedents/ consequences as the construct</td>
<td>Measures may not have the similar significance of relationships with the antecedents/consequences as the construct</td>
</tr>
<tr>
<td>Taking measurement error into account at the measurement level; error terms in indicators can be identified</td>
<td>Taking measurement error into account at the construct level; error term cannot be identified if the formative measurement model is estimated in isolation</td>
</tr>
<tr>
<td>Assessments: Internal consistency reliability, Indicator reliability, Convergent validity, Discriminant validity Fornell-Larcker criterion, Cross loadings, and Heterotrait-monotrait ratio of correlations (HTMT)</td>
<td>Assessments: Multicollinearity, Construct Validity, and Indicator Reliability</td>
</tr>
</tbody>
</table>
As per Table 6.1, the most typical distinction between reflective and formative measures has to do with the relationship between the construct and its measurement items. The reflective modes indicate causality from constructs to measurement items, whereas formative modes reflect the opposite. In formative measurement models, the latent variable is considered a consequence of its respective indicators and because the latent variable is defined by its indicators, changing/replacing a formative indicator will alter the meaning of the latent variable. Alternatively, in reflective measurement models, indicators are regarded as the consequences of the latent variable to which they belong, which means items are manifested by the construct. The reflective indicators can be used interchangeably and can, to a certain extent, even be discarded. Another critical differentiation between the two models has to do with whether or not the measurement items are correlated. In the formative model, it is not essential for all measurement items to be highly correlated, while the reflective model stipulates that all measurement items need to have a high level of correlation.

7. CONCLUSION

This paper proposes a set of guidelines in classifying the formative and reflective constructs and the steps in assessing the formative and reflective measurement model. In addition, this paper confirms that there are apparent differences between reflective and formative constructs and the construct identification and validation depends on the type of construct specified by the researcher. This paper proposes that quantitative researchers that the decision whether to use a formative or reflective indicator should be based on the theoretical grounds. Misspecification of measurement models may affect research outcome or mislead future research.

Conflicts of Interest: The author(s) declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgement: This research work was supported by the Universiti Teknologi MARA Malaysia.
References


