

# 4

## ASSESSING PLS-SEM RESULTS PART I

### Evaluation of Reflective Measurement Models

#### CASE STUDY ILLUSTRATION—EVALUATION OF REFLECTIVE MEASUREMENT MODELS

---

##### **Running the PLS-SEM Algorithm**

We continue working with our PLS-SEM example on corporate reputation. In Chapter 3, we explained how to estimate the PLS path model and how to obtain the results by opening the default report in the SmartPLS 4 software (Ringle, Wende, & Becker, 2022). Recall that to do so, you must first load the simple corporate reputation model and then run the model by clicking on the wheel symbol with the label *Calculate* in the toolbar or by using the pull-down menu by going to *Calculate* → *PLS-SEM algorithm*. After running the *PLS-SEM algorithm*, if you had previously ticked *Open report* on the dialog box, the SmartPLS results report automatically opens; if not, click on the *Open Report* button in the *Modeling window*. You can save the results of this report by clicking on the *Save* button in the toolbar.

Before analyzing the results, you need to quickly check if the algorithm converged (i.e., the stop criterion of the algorithm was reached and not the maximum number of iterations). To do so, go to *Algorithm* → *Stop criterion changes* in the results report. You will then see the table shown in Exhibit A4.1, which shows the number of iterations of the PLS-SEM algorithm. This number should be lower than the maximum number of iterations (i.e.,

3,000). At the bottom left side of the table, you will see that the algorithm converged after Iteration 5. If the PLS-SEM algorithm does not converge in fewer than 3,000 iterations (the default setting in the software), the algorithm could not find a stable solution. This kind of situation almost never occurs. But if it does occur, there are problems with the data and they need to be checked carefully. For example, data problems may occur if the sample size is too small or if an indicator has many identical values (i.e., the same data points, which results in insufficient variability).

**Exhibit A4.1 ■ Stop Criterion Table in SmartPLS**

	comp_1	comp_2	comp_3	cusa	cusl_1	cusl_2	cusl_3	like_1	like_2	like_3
Iteration 0	0.401	0.401	0.401	1.000	0.386	0.386	0.386	0.386	0.386	0.386
Iteration 1	0.536	0.341	0.328	1.000	0.368	0.421	0.365	0.419	0.378	0.359
Iteration 2	0.536	0.340	0.328	1.000	0.369	0.420	0.365	0.418	0.378	0.360
Iteration 3	0.536	0.340	0.328	1.000	0.369	0.420	0.365	0.418	0.378	0.360
Iteration 4	0.536	0.340	0.328	1.000	0.369	0.420	0.365	0.418	0.378	0.360
Iteration 5	0.536	0.340	0.328	1.000	0.369	0.420	0.365	0.418	0.378	0.360

When your PLS path model estimation converges, which it practically always does, you need to examine the following PLS-SEM calculation results tables from the results report for reflective measurement model assessment: *Outer loadings*, *Composite reliability (rho\_a)*, *Composite reliability (rho\_c)*, *Cronbach's alpha*, *Average variance extracted (AVE)*, and *Discriminant validity*. We examine other elements of the report in Chapters 5 and 6, when we extend the simple path model by including formative measures and also consider the structural model results.

**Reflective Measurement Model Evaluation**

The simple corporate reputation model has three latent variables with reflective measurement models (i.e., *COMP*, *CUSL*, and *LIKE*) as well as a single-item construct (*CUSA*). For the reflective measurement models, we need the estimates for the relationships between the reflective constructs and their indicators (i.e., outer loadings). Exhibit A4.2 displays the results table for the outer loadings, which can be found under *Final results* → *Outer loadings*. By default, the outer loadings are also displayed in the *Graphical output* of the PLS results report after running the PLS-SEM algorithm. All outer loadings of the reflective constructs *COMP*, *CUSL*, and *LIKE* are well above the threshold value of 0.708, which suggests sufficient levels of indicator reliability.

**Exhibit A4.2 ■ Outer Loadings**

	COMP	CUSA	CUSL	LIKE
comp_1	0.858			
comp_2	0.798			
comp_3	0.818			
cusa		1.000		
cusl_1			0.833	
cusl_2			0.917	
cusl_3			0.843	
like_1				0.879
like_2				0.870
like_3				0.843

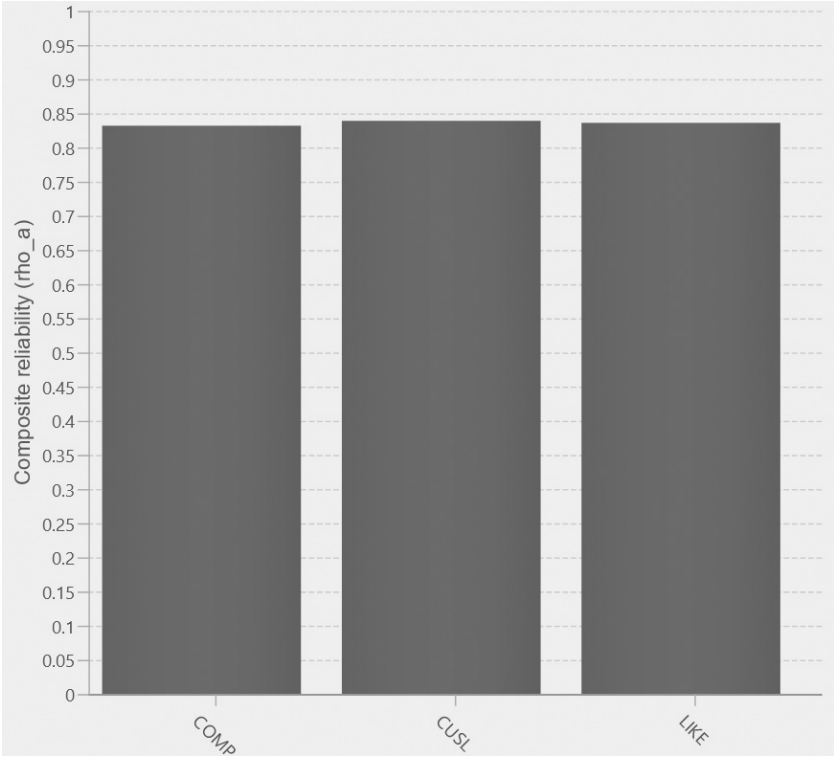
The indicator *comp\_2* (outer loading: 0.798) has the smallest indicator reliability with a value of 0.637 ( $0.798^2$ ), while the indicator *cusl\_2* (outer loading: 0.917) has the highest indicator reliability, with a value of 0.841 ( $0.917^2$ ).

Next, go to *Quality criteria* → *Construct reliability and validity* to evaluate the construct measures' internal consistency reliability and convergent validity. Here, you have the option of displaying the composite reliability values using a bar chart or in a matrix format. Exhibit A4.3 shows the internal consistency reliability values in matrix format. With  $\rho_A$  values of 0.832 (*COMP*), 0.839 (*CUSL*), and 0.836 (*LIKE*), all three reflective constructs have high levels of internal consistency reliability. Clicking on *Composite reliability (rho\_a)*—*Bar chart* shows the bar chart of the constructs' reliability values (Exhibit A4.4). If a  $\rho_A$  value is above the threshold value, the corresponding bar is colored green. If a  $\rho_A$  is lower than 0.70, the bar is colored red. As indicated above, all  $\rho_A$  values exceed the threshold. Note that the  $\rho_A$  value of the single-item variable *CUSA* is 1.00, but this cannot be interpreted as evidence the construct measurement is perfectly reliable and should not be reported with other measures of reliability.

**Exhibit A4.3 ■ Construct Reliability and Validity**

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
COMP	0.776	0.832	0.865	0.681
CUSL	0.831	0.839	0.899	0.748
LIKE	0.831	0.836	0.899	0.747

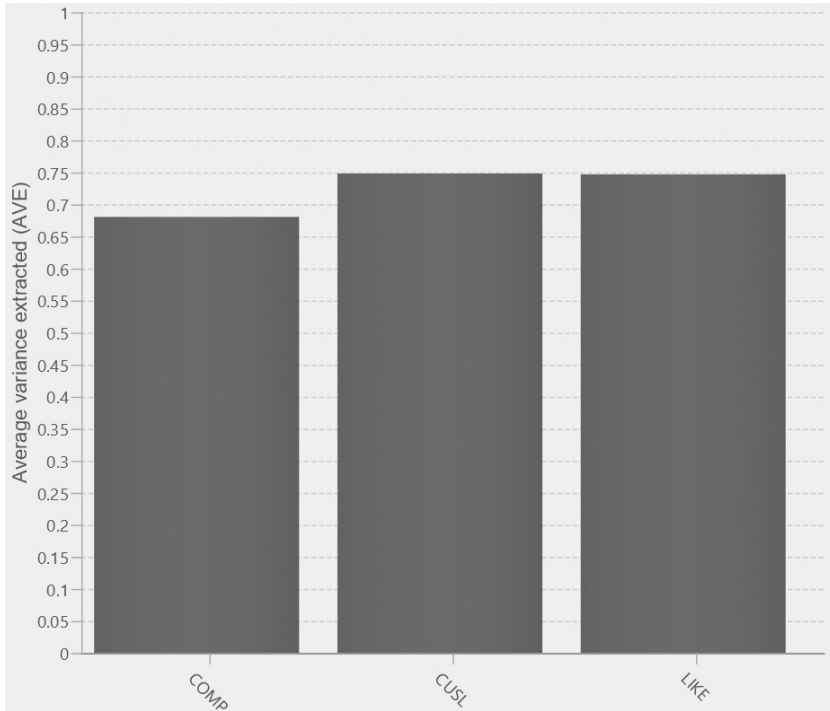
**Exhibit A4.4 ■ Reliability  $\rho_A$**



Analogous to  $\rho_A$ , you can also examine the bar charts of Cronbach’s alpha and composite reliability  $\rho_C$ . All bars in the charts appear in green, indicating that all construct measures are above the 0.70 threshold. The specific values of Cronbach’s alpha (0.776 for *COMP*, 0.831 for *CUSL*, and 0.831 for *LIKE*) and the composite reliability  $\rho_C$  (0.865 for *COMP*, 0.899 for *CUSL*, and 0.899 for *LIKE*) can be accessed by going to *Quality criteria* → *Construct reliability and validity* → *Overview*. Again, as *CUSA* is measured using a single item, interpreting this construct’s Cronbach’s alpha or composite reliability values is not meaningful.

Convergent validity assessment is based on the AVE values, which can be accessed by navigating to *Quality criteria* → *Construct reliability and validity* in the *Results report*. As with the internal consistency reliability measures, SmartPLS offers the option of displaying the results using bar charts (Exhibit A4.5) or in a matrix format. In this example, the AVE values of *COMP* (0.681), *CUSL* (0.748), and *LIKE* (0.747) are well above the required mini-

### Exhibit A4.5 ■ Average Variance Extracted (AVE)



imum level of 0.50. Thus, the measures of the three reflective constructs have high levels of convergent validity.

Finally, in the *Quality criteria* → *Discriminant validity*, SmartPLS offers several approaches to assess whether the construct measures empirically demonstrate discriminant validity. The primary criterion for discriminant validity assessment is the *Heterotrait-monotrait ratio (HTMT)*, which can be accessed via the *Quality criteria* → *Discriminant validity* → *Heterotrait-monotrait ratio (HTMT)—Matrix* section of the results report. Exhibit A4.6 shows the HTMT values for all pairs of constructs in a matrix format. Note that SmartPLS computes the HTMT values on the grounds of absolute correlations to avoid adverse results due to a combination of positive and negative indicator correlations (Ringle, Sarstedt, Sinkovics, & Sinkovics, 2023). The *Results report* also shows these HTMT values in a list format and in bar charts, using 0.85 as the relevant threshold level. As can be seen, all HTMT values are clearly lower than the more conservative threshold value of 0.85, even for *COMP* and *LIKE* as well as *CUSA* and *CUSL*, which are very similar from a conceptual viewpoint. Recall that the threshold value for conceptually similar constructs is 0.90.

According to the Fornell-Larcker criterion, the square root of the AVE of each construct should be higher than the construct’s highest correlation with any other construct in the model (this notion is identical to comparing the AVE with the squared correlations between the constructs). Exhibit A4.7 shows the results of the Fornell-Larcker criterion assessment with the square root of the reflective constructs’ AVE on the diagonal and the correlations between the constructs in the off-diagonal position. For example, the reflective construct *COMP* has a value of 0.825 for the square root of its AVE, which needs to be compared with all correlation values in the column of *COMP*. Note that for *CUSL*, you need to consider the correlations in both the row and column and for *LIKE* only those in the row. Overall, the square roots of the AVEs for the reflective constructs *COMP* (0.825), *CUSL* (0.865), and *LIKE* (0.864) are all higher than the correlations of these constructs with other latent variables in the path model, thus indicating all constructs are valid measures of unique concepts.

Note that while frequently used in past applied research, the Fornell-Larcker criterion does not reliably detect discriminant validity issues. That is, the criterion is not well-suited for disclosing existing discriminant validity problems. Nevertheless, any violation of the Fornell-Larcker criterion should be taken as strong evidence for a severe discriminant validity problem.

In addition to examining the HTMT ratios, you should test whether the HTMT values are significantly different from the threshold value.

**Exhibit A4.6 ■ HTMT**

	COMP	CUSA	CUSL	LIKE
COMP				
CUSA	0.465			
CUSL	0.532	0.755		
LIKE	0.780	0.577	0.737	

**Exhibit A4.7 ■ Fornell-Larcker Criterion**

	COMP	CUSA	CUSL	LIKE
COMP	0.825			
CUSA	0.436	1.000		
CUSL	0.450	0.689	0.865	
LIKE	0.645	0.528	0.615	0.864

Specifically, we assume a 0.85 threshold for all pairs of constructs except for *COMP* and *LIKE* as well as *CUSA* and *CUSL*, for which we assume a higher threshold (0.90) because of their conceptual similarity. To run this test, we need to compute bootstrap confidence intervals using bootstrapping. To run the bootstrapping procedure, go back to the *Modeling* window by clicking on *Edit* in the *Results window*. Next, left-click on *Calculate* → *Bootstrapping* in the pull-down menu. In the dialog box that opens, choose the bootstrapping options as displayed in Exhibit A4.8 (Chapter 5 includes a more detailed introduction to the bootstrapping procedure and the parameter settings). Make sure to select 10,000 *Subsamples* and the *Complete (slower)* option, which includes the results for HTMT (unlike the *Most important (faster)* option). Select *Percentile bootstrap* as the *Confidence interval method*, *One tailed* as *Test type*, and a 0.05 *Significance level* (note that you can also select *Two tailed* as *Test type* and a 0.10 *Significance level*). Note that *Fixed seed* next to *Random number generator* should be selected to facilitate the reproducibility of results. Finally, make sure to check the box next *Open report* and click on *Start calculation*.

**Exhibit A4.8 ■ Bootstrapping Options in SmartPLS**

The screenshot shows the 'BT setup' dialog box in SmartPLS. The settings are as follows:

- Subsamples:** 10000
- Do parallel processing
- Amount of results:** Complete (slower)
- Confidence interval method:** Percentile bootstrap
- Test type:** One tailed
- Significance level:** 0.05
- Random number generator:** Fixed seed

At the bottom of the dialog, there is a checked box for **Open report**. Navigation buttons include **Default settings** (link), **Close** (button), and **Start calculation** (button).

After running bootstrapping, the results report opens automatically. Go to *Quality criteria* → *Heterotrait-monotrait ratio (HTMT)* and navigate to *Confidence intervals bias corrected*. The table that opens (Exhibit A4.9) shows the original HTMT values (column *Original sample (O)*) for each combination of constructs in the model, along with the average HTMT values computed from the 10,000 bootstrap samples, as shown in column *Sample mean (M)*. Note that you should have the same results, provided that you selected the *Fixed seed* option in the bootstrapping settings. However, if you chose *Random seed*, your results will differ from those presented here and will change slightly each time you rerun the bootstrapping procedure. The reason is that bootstrapping builds on randomly drawn bootstrap samples, which will differ every time the procedure is run. The differences in the overall bootstrapping results are marginal, however, provided that a sufficiently large number of bootstrap samples have been drawn (e.g., 10,000). By selecting the *Fixed seed* option, the bootstrapping algorithm's random number generator uses fixed settings, which facilitates reproducing the results in later model estimations.

**Exhibit A4.9 ■ Percentile Confidence Intervals (with Bias Correction) for HTMT**

	Original sample (O)	Sample mean (M)	Bias	5.0%	95.0%
CUSA <-> COMP	0.465	0.464	-0.001	0.369	0.556
CUSL <-> COMP	0.532	0.532	-0.000	0.429	0.625
CUSL <-> CUSA	0.755	0.754	-0.000	0.695	0.805
LIKE <-> COMP	0.780	0.780	0.000	0.705	0.843
LIKE <-> CUSA	0.577	0.577	-0.000	0.503	0.644
LIKE <-> CUSL	0.737	0.737	-0.000	0.664	0.800

The columns labeled *5%* and *95%* show the lower and upper bounds of the 95% one-sided bootstrap confidence interval (or the 90% two-sided bootstrap confidence interval, respectively). The statistical test focuses on the right tail of the bootstrap distribution to show that HTMT value is significantly lower than the corresponding threshold values (0.85 and 0.90) with a 5% probability of error. As can be seen all values in the *95%* column are lower than these threshold values, suggesting that all HTMT values are significantly lower than 0.85 and 0.90, respectively. More importantly, even assuming a more conservative threshold of 0.85 for all construct combinations (i.e., including *CUSA* and *CUSL* as well as *COMP* and *LIKE*), we find that all HTMT values are significantly lower than this value (i.e., the confidence intervals' upper bound is smaller than 0.85). For example, the lower and upper bounds of the HTMT's 95% confidence interval for *CUSA* and *COMP* are



**Exhibit 4.10 ■ Results Summary for Reflective Measurement Models**

Latent Variable	Indicators	Convergent Validity			Internal Consistency Reliability			Discriminant Validity
		Loadings	Indicator Reliability	AVE	Cronbach's Alpha	Reliability $\rho_A$	Composite Reliability $\rho_C$	
		>0.70	>0.50	>0.50	0.60–0.90	0.60–0.90	0.60–0.90	
COMP	<i>comp_1</i>	0.858	0.736	0.681	0.776	0.832	0.865	Significantly lower than 0.85 (0.90)? Yes
	<i>comp_2</i>	0.798	0.637					
	<i>comp_3</i>	0.818	0.669					
CUSL	<i>cusl_1</i>	0.833	0.694	0.748	0.831	0.839	0.899	Yes
	<i>cusl_2</i>	0.917	0.841					
	<i>cusl_3</i>	0.843	0.711					
LIKE	<i>like_1</i>	0.879	0.773	0.747	0.831	0.836	0.899	Yes
	<i>like_2</i>	0.870	0.757					
	<i>like_3</i>	0.843	0.711					

0.369 and 0.556, respectively. Since the upper bound of 0.556 is lower than 0.85, the HTMT value of 0.465 for *CUSA* and *COMP* is significantly lower than the more conservative threshold value of 0.85. To summarize, the bootstrap confidence interval results of the HTMT criterion also clearly demonstrate the discriminant validity of the constructs.

Exhibit A4.10 summarizes the results of the reflective measurement model assessment. As can be seen, all model evaluation criteria have been met, providing support for the measures' reliability and validity.