MARKETING OR METHODOLOGY? EXPOSING THE FALLACIES OF PLS WITH SIMPLE DEMONSTRATIONS

Journal:	*European Journal of Marketing*
Manuscript ID	EJM-02-2021-0099.R3
Manuscript Type:	Original Article
Keywords:	PLS, Partial Least Squares, Measurement, Composites, Theory Testing, Structural Equation Models, Model Testing

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

capitalization on chance simulations.R
Corporate Reputation demo.R
ECSI demo.R
TAM demo.R

MARKETING OR METHODOLOGY? EXPOSING THE FALLACIES OF PLS WITH SIMPLE DEMONSTRATIONS

STRUCTURED ABSTRACT

PURPOSE

Over the last twenty years, PLS has become a popular method in marketing research. At the same time, several methodological articles have demonstrated problems with the technique, but have had little impact on its use in marketing research practice. This article aims to present some of these criticisms in a reader-friendly way for non-methodologists.

DESIGN/METHODOLOGY/APPROACH

Key critiques of PLS are summarized and demonstrated using existing datasets in easily replicated ways. Recommendations are made for assessing whether PLS is a useful method for a given research problem.

FINDINGS

PLS is fundamentally just a way of constructing scale scores for regression. PLS provides no clear benefits for marketing researchers and has disadvantages that are features of the original design and cannot be solved within the PLS framework itself. Unweighted sums of item scores provide a more robust way of creating scale scores.

ORIGINALITY

This work presents a novel perspective on PLS critiques by showing how researchers can use their own data to assess whether PLS (or another composite method) can provide any advantage over simple sum scores. A Composite Equivalence Index (CEI) is introduced for this purpose.

KEYWORDS: Partial Least Squares, Measurement, Composites, Structural Equation Models

"Most people use statistics like a drunk man uses a lamppost; more for support than illumination" --- Andrew Lang, Scottish Novelist

Partial least squares (PLS) is an algorithm developed by Herman Wold in the 1960s and 70s (Wold, 1982) and was originally positioned as an alternative to the LISREL program (Jöreskog & Wold, 1982, Equations 4-8). The main stated advantage at that time was that the "PLS approach is easy and speedy on the computer" (Wold, 1982, p. 29), but this came with the trade-off that PLS produces incorrect estimates of model parameters: "LISREL gives consistent estimates of the structural parameters, whereas the corresponding PLS estimates are biased" (Wold, 1982, p. 52). Given advances in computing power since Wold's work, any advantage PLS may hold in computational simplicity is moot. However, the disadvantages remain.

Despite early exposure in the marketing discipline (Fornell & Bookstein, 1982), PLS remained a niche method until the early 2000s when its popularity began to increase. This trend accelerated from around 2010 (Hair et al., 2014, fig. I.1; Rönkkö, 2014, fig. 1), driven by advocacy papers with titles like *"PLS-SEM: Indeed a Silver Bullet"* (Hair et al., 2011). At the same time, several critiques of PLS appeared (e.g., Goodhue et al., 2012; Rönkkö & Evermann, 2013) but have had little impact on PLS's use in marketing journals, for three possible reasons. First, some critiques of PLS are in methodological journals (e.g., Rönkkö & Evermann, 2013) that applied researchers may not follow. Indeed, a researcher following only marketing journals might have read at least half-a-dozen advocacy papers but not a single critical one. Second, there is a widespread belief that the critical arguments have been refuted. For example, Ali, Mostafa, and Cobanoglu (2018) claim that "most of the criticism has been refuted as inaccurate" (p. xi), while Ravand and Baghaie (2016) state that "Henseler et al. (2014) refuted the critiques of Rönkkö and Evermann" (p. 3). It is difficult to see what these conclusions are based on because the evidence presented by Henseler et al. (2014)

mostly *supported* Rönkkö and Evermann's (2013) arguments (McIntosh et al., 2014). Third, researchers may conclude that because more papers champion than critique PLS, it must be valid. This is a logical fallacy because the number of advocacy articles is not evidence of PLS's usefulness; it simply shows that the advocates are more prolific writers than the skeptics.

This article presents a few key methodological criticisms of PLS using simple examples that any reader capable of using PLS can replicate on publicly available data or their own datasets. For each claim, the arguments presented in literature advocating PLS are summarized, and their invalidity is demonstrated. A new metric is proposed for researchers to assess whether PLS – or any other indicator weighting system – can make a difference in a given situation.

WHAT PLS DOES AND WHY IT IS PROBLEMATIC

Introductory texts (e.g., Hair et al., 2014) describe PLS as a structural equation model (SEM) estimation technique that is compared to the maximum likelihood estimation of SEMs with latent variables (ML-SEM). These techniques are often presented as "second-generation" techniques that are claimed to be *a-priori* superior to regression, exploratory factor analysis, or ANOVA, which are presented as "first-generation" techniques as shown in Panel A of Figure 1. However, this classification is based simply on when the techniques were introduced to the marketing discipline (see Fornell, 1987), rather than any methodological principle suggesting that "second-generation" methods are superior to "first-generation" methods.

Methods should be chosen based on their characteristics, instead of when they were introduced to a field. With multiple-item data, the most fundamental decision is whether to aggregate the data as scale scores or use them to estimate a latent variable model. Although latent variable models are often considered superior because they can model measurement error, this advantage rests on the correct measurement model specification. Unfortunately, incorrect measurement models may cause larger bias than simply using scale scores (Rhemtulla et al., 2020), complicating the choice between these approaches.

After deciding between latent variables and scale scores, more specific choices are needed as shown in Panel B of Figure 1. When using scale scores, researchers often default to linear composites (i.e., weighted sums) for simplicity, leaving only the choice of weighting system, of which PLS is one alternative. Unfortunately, researchers are ill-served by the existing literature regarding both awareness of these choices and guidance in making them. Most PLS articles obscure the fact that PLS is not a latent-variable method at all, but an indicator weighting system that creates composite scores for subsequent regression analysis (Evermann & Rönkkö, 2021; Goodhue et al., 2012; Rönkkö & Ylitalo, 2010). In fact, the indicator weighting is the *only* difference between PLS and using regression with scale scores calculated as sums or means of items, which many researchers learn as a first technique for analyzing multiple-item data.

------ Figure 1 ------

Because one of the best ways to communicate an idea is to help a person to demonstrate it to themselves, this paper illustrates PLS using three publicly-available datasets, allowing readers to replicate these examples on their own and in the classroom. The data are the ECSI dataset from Tenenhaus et al. (2005), the “corporate reputation” example from Hair et al. (2014, Chapter 2), and the TAM dataset from SmartPLS (2020). These datasets were not chosen to obtain a particular result; rather, they were chosen for availability and potential familiarity to the reader. Figure 2 shows the path diagrams for the ECSI and corporate reputation models. (The TAM model is omitted because this dataset is not discussed in detail in the article). The analyses uses R but the online supplement provides screencasts for replication with SPSS and SmartPLS.

------ Figure 2 ------

FALLACY 1: PLS MAXIMIZES EXPLAINED VARIANCE OR R^2

The PLS textbook by Hair et al. (2014) starts explaining PLS by stating that "PLS-SEM estimates coefficients (i.e., path model relationships) that maximize the R^2 values of the (target) endogenous constructs" (p. 14). This claim is never explained but is repeated throughout the book, making it *appear* important. The same is also stated in some PLS criticisms (Goodhue et al., 2012, p. 984) and countless empirical applications, making it important to address.

The R^2 maximization claim is a variant of a more general claim that the PLS weights are somehow optimal (e.g., Chin, 1998, p. 307; Henseler & Sarstedt, 2013, p. 566). These optimality claims are typically vague, lacking explanations for what purpose the weights would be optimal for, and are evidence-free, lacking any proofs of optimality. The specific R^2 maximization claim is imprecise as there may be multiple R^2 values in a model, and it is unclear which function of R^2 is maximized (mean, sum of squares, etc.), and because PLS is a combination of multiple inner and outer estimation algorithms, and it is unclear which combination produces the maximum. It is also unclear why R^2 maximization would be useful for estimating parameters in a complex model consisting of multiple equations.

Empirical demonstration that PLS does not maximize R^2

To show that PLS does not maximize R^2, it is sufficient to show that another technique produces a larger R^2 value. Indeed, Rönkkö (2020a, sec. 2.3) calculates indicator weights to explicitly optimize R^2, leading to an R^2 value more than double that produced by PLS. The same can be done with any empirical dataset. First, specify a model with one dependent composite. To illustrate, Loyalty is predicted by Image, Satisfaction, and Complaints using the ECSI data. Next, run this model using PLS. For comparison, run a canonical correlation analysis with Image, Satisfaction, and Complaints indicators as x variables and Loyalty indicators as y variables. The results are shown in Table 2. PLS Mode A does not perform well in maximizing R^2. Mode B is better but still produces a smaller R^2 than canonical correlation weights do.

------ Table 2 ------

Conclusions on R^2 maximization

If the objective is to maximize R^2, PLS is demonstrably the wrong choice. If maximization of a single R^2 value (or any other statistic) was of interest, the first step should be to define a clear maximization objective. Then, the maximization problem could be solved by a general optimization routine, or by a problem-specific algorithm. Instead, PLS seems to be a solution in search of a problem.

FALLACY 2: PLS WEIGHTS IMPROVE RELIABILITY

The most important question about PLS is whether it produces better composites than the alternatives. That is, after a researcher has chosen to use a) scale scores instead of latent variables, and b) linear composites as the calculation strategy (see Figure 1), she needs to ask whether c) PLS weights are somehow better than, for example, equal or unit weights[1]. Of the many explanations regarding the potential advantages of PLS weights (Rönkkö, McIntosh, Antonakis, et al., 2016), the assertion that PLS weights maximize reliability is the most popular. For example, Gefen, Ridgon, and Straub (2011) state that: "Optimization of [the] weights aims to maximize the explained variance of dependent variables. [...] maximizing explained variance will also tend to minimize the presence of random measurement error in these latent variable proxies" (p. v), and Hair et al. (2014) note that PLS "prioritizes the indicators according to their individual reliability" (p. 101). However, no evidence has been reported to support these claims.

[1] Unit weights refer to applying equal weights after standardization. These terms are used interchangeably in this article because standardization is used in all examples.

Decades of literature show that as far as reliability is concerned "there is overwhelming evidence that the use of differential weights [over unit weights] seldom makes an important difference" (Nunnally, 1978, p. 297; see Wang & Stanley, 1970 for a review). Because empirically-determined weights can provide only marginal advantages and may have serious drawbacks, the usual recommendation is to use unweighted composites (Cohen, 1990; Graefe, 2015; Grice, 2001). Indeed, Rönkkö and Ylitalo (2010) demonstrated that PLS weights can harm reliability and validity (see Rönkkö & Evermann, 2013). Henseler et al (2014) objected to these conclusions, but their simulations demonstrated only trivial advantages of PLS weights over unit weights – only a 0.6% increase in reliability in favorable situations – and a serious *loss* in reliability of 16.8% in less favorable scenarios (Henseler et al, 2014, Table 2). Recently Henseler (2021) appears to concede the superiority of unit weights, writing that "Sum scores can be a good choice [...] Particularly if the observed variables are highly correlated, [...] differential weighting hardly excels over sum scores" (p. 87). The simulations by Rönkkö et al. (2016) suggest that inter-item correlations of .4 or greater are sufficient to eliminate the effects of PLS weights on reliability even in otherwise ideal conditions.

Gefen, Rigdon, and Straub (2011) provide another perspective on indicator weights, stating that the "weights of the measurement items associated with the same latent variable should be approximately the same, unless researchers have *a-priori* theory-based expectations of substantial differences in performance across items" (p. viii). This leaves little room for the PLS weights because on the one hand, if weights are not suggested by a theory, using equal weights is a simpler and more robust solution, and on the other hand, if a theory suggests a set of weights, that set should be used instead of empirically determined weights (Lee et al., 2013).

Empirical demonstration that PLS does not improve reliability

If PLS composites were to provide any advantage over unweighted composites, i.e. simple item sums, they should at least differ from them. Yet, Table 1 shows that when the ECSI model in Figure 2 is estimated, the two kinds of composites are nearly indistinguishable, correlating perfectly at two-digit precision. The correlation between Loyalty composites at .93 is an exception. To understand whether the PLS composite is better than the unweighted one, it is essential to check the weights and to understand why they differ. In this case, the PLS weights are .45, .13, and .66, showing that PLS downweighted the second indicator. Factor analysis of the Loyalty scale produces a loading of .10 for the second indicator. Normally, an indicator with such a low loading would be dropped and this is what Tenenhaus et al. (2005) did. After this, the unweighted and the PLS composites correlate at .99 and thus one cannot have a meaningful advantage over the other.

------ Table 1 ------

Repeating the same analysis with the other two datasets did not produce a single correlation below the .99 level (online supplement). Rönkkö et al. (Rönkkö et al., 2015; Rönkkö, McIntosh, Antonakis, et al., 2016) show similar results with two additional datasets, establishing a clear pattern: PLS weights do not appear to provide any advantages for data that they are commonly applied to.

Conclusions on reliability improvement

The comparisons of PLS and unweighted composites show what has been known for decades: differential weights rarely make a difference. As long as the indicators are at least moderately correlated, advantages from weights are trivial (e.g., Cohen, 1990; Graefe, 2015; Grice, 2001). Nevertheless, a few recent articles have presented simulations where PLS weights make a difference (e.g., Becker et al., 2013; Hair et al., 2017). These studies appear not to be designed to be representative of real datasets but simply to find scenarios where indicator weights make a maximal difference. For example, Becker et al. (2013) used four-variable scales consisting of two uncorrelated pairs. It is difficult to imagine what kind of measurement process would produce such data[2], and none of the empirical datasets used for demonstrating PLS have this kind of correlational pattern.

Although the idea that weighted composites may have advantages over unweighted composites is intuitively appealing, there is clear evidence that such benefits do not exist in practice. As Cohen (1990) puts it: "as a practical matter, most of the time, we are better off using unit weights: +1 for positively related predictors, -1 for negatively related predictors, and 0, that is, throw away poorly related predictors" (p. 1306). Considering the potential *disadvantages* of PLS, covered next, improving reliability is certainly not a reason to use PLS.

UNTOLD FACT: PLS WEIGHTS CAN BIAS CORRELATIONS

There are two important cases where PLS composites *do* differ from unweighted composites but in a negative way. First, when two scales are only weakly correlated, PLS can inflate regression coefficients. Second, if there are cross-loadings or correlated errors between different scales, PLS tends to inflate the resulting biases. Consider the simple model in Figure 3. In this case, PLS weights the a indicators by their correlations with the b indicators (Rönkkö, 2014). The population correlations are equal at 0.147, and when applied to population data, PLS produces equal weights as seen in Table 3. If for some reason a_1 correlated more strongly with the b indicators than a_2 or a_3 do, a_1 would receive a higher weight than a_2 or a_3. We demonstrate these effects by increasing one correlation by 0.1 and decreasing another one by 0.1, marked by dashed lines in Figure 3. As shown in Table 3, PLS

[2] Proponents of "formative measurement" state that formative indicators do not need to be correlated. Even so, they generally are at least moderately correlated in practice.

weights the indicators with the positive error correlation higher and those with the negative correlation lower, producing a 6% larger correlation between composites.

------ Figure 3, Table 3 ------

The effects of chance correlations on PLS weights

Recent studies have shown that PLS capitalizes on chance in small samples (Goodhue et al., 2015; Rönkkö, 2014; Rönkkö & Evermann, 2013). In sample data, the correlations between the *a* and *b* indicators vary around their population values simply by chance. With a sample of 100, the standard deviation of the correlation is 0.100, making the inflation of composite correlations likely in any given sample, i.e. PLS capitalizes on chance correlations. The magnitude of the bias depends on the strength of the latent variable correlations. Figure 4 shows the results of applying different analysis techniques to 1000 simulated samples (N=100) from Figure 3, varying the latent variable correlation. The differences are clear: Both sets of PLS results are biased away from zero, producing a small secondary peak (mode) of negative estimates. As the population correlation increases the PLS estimates approach equal-weight estimates. In all cases, ML-SEM estimates are unbiased.

------ Figure 4 ------

Capitalization on chance explains the simulation results by Chin and Newstedt (1999), which were pivotal in starting the myth that PLS would be particularly appropriate for small samples (Rönkkö, 2014; Rönkkö & Evermann, 2013). This bias is also *solely* due to sampling error and could be *completely avoided with no downsides* by using equal weights. Indeed, when discussing these findings Henseler (2021) agrees that "sum scores are a viable approach to mitigate problems of 'chance correlations' as described by Rönkkö (2014)" (p. 87). However, because it is difficult to know *a-priori* if constructs are highly correlated (and estimating these is surely a key purpose of a typical research study), using equal weights is *always* a better choice in real research situations.

Empirical demonstration of bias due to chance correlations

Because none of the example datasets contain any weakly correlated scales, the problem of chance correlations is demonstrated here by generating additional variables for the ECSI dataset. A six-sided die was rolled (simulated in R) and the values were recorded into four new variables (Latent die, Error die 1, Error die 2, and Error die 3), with three new variables die1, die2, and die3 created as sums of the Latent die and each of the Error die variables. Thus, the die variables form a scale that is uncorrelated with the other scales.

PLS analysis was run in three different configurations shown in Table 4, using a subset of the variables for simplicity. The table provides three key takeaways: First, the PLS weights, particularly for Loyalty and Satisfaction, vary widely from one analysis to the next. Second, the correlations involving the Die composite are always stronger when using PLS weights than when using unit weights due to capitalization on chance. Third, the correlations between PLS composites are always larger when the correlation corresponds to a regression path (are "adjacent"). For empirical demonstrations of this PLS feature, see Rönkkö, McIntosh, et al. (2016, Table 2). The capitalization on chance effect can be seen in the distribution of the bootstrap replications of the regression estimates shown in Figure 5.

------ Table 4 and Figure 5 -------

Empirical demonstration of the effects of cross-loadings on PLS weights

The effect of cross-loadings is demonstrated using the corporate reputation data from Hair et al. (2014, Chapter 2). Factor analysis results in Table 5 show that indicator Comp1 cross-loads strongly on the Comp and Like factors and the Cusl1 indicator has a weaker cross-loading on Like. Recommended research practice is to omit the problematic indicator (Hair, 2010, Chapter 3), but Table 6 shows that PLS does the exact opposite, assigning Comp1 indicator the highest weight. This overweighting of Comp1 causes the Comp composite to contain more variance of Like than it should, and the regression coefficient of the Comp composite is increased at the cost of decreasing the coefficient of the Like composite.

------ Table 5, Table 6 -------

In the previous example, the cross-loading affected two non-adjacent scales. To demonstrate the effects of a cross-loading between two adjacent scales, one was artificially created between Like2 and Cusl by calculating a new variable as $Like2_{new} = Like2 + \frac{Cusl1 + Cusl2 + Cusl3}{3}$. The second set of columns of Table 6 shows the results from rerunning the analyses using these manipulated data. For unit weights, the cross-loading inflates the regression coefficient by 60% from .333 to .535. For PLS weights, the coefficient is inflated by 80% from .331 to .599. Thus, PLS exacerbates the problem of cross-loadings. Unfortunately, PLS is often applied without following the strict recommendation that one should "never create a [composite] without first assessing its unidimensionality with exploratory or confirmatory factor analysis" (Hair, 2010, p. 127), causing problems like the ones shown above to easily escape detection.

Conclusions on bias due to PLS weights

With correlated indicators, indicator weights rarely make a difference. Two known scenarios where PLS weights *do* make a difference are capitalization on chance when indicators are only weakly correlated across scales, and inflating the effects of cross-loadings. Rigdon (2016) claims that weakly correlated scales present a well-known violation of the assumptions of PLS. Unfortunately, except for the new book by Henseler (2021), not a single introductory text makes this assumption explicit. Further, it is unclear how a researcher could know that their scales are weakly correlated in advance, nor is it clear why a method that cannot deal with weakly correlated scales would be of any use in real research situations. Fortunately, these problems are easily avoided by using equal weights in the analysis.

NEW METHODOLOGICAL PROPOSAL: COMPOSITE EQUIVALENCE INDEX (CEI)

The previous sections demonstrated that using PLS does not improve reliability meaningfully and can lead to problems in small samples or for cross-loading items. Hence, researchers should always consider unweighted composites as the first choice and "always include this simple contender and test more sophisticated alternatives against it" (Dijkstra, 2009, p. 5).

The *Composite Equivalence Index* (CEI) is proposed to determine if the PLS composites differ substantially from unweighted composites. The CEI can be calculated by exporting the composites from PLS software and correlating these with unit-weighted composites. Two variants of the index are proposed. $\text{CEI}_{\text{individual}}$ is the correlation of each PLS composite with the corresponding unweighted composite; $\text{CEI}_{\text{minimum}}$ is the minimum of the $\text{CEI}_{\text{individual}}$ and quantifies whether PLS weights make a difference at all for the analysis.

The CEI statistic can be expressed in matrix form:

$$CEI = diag\left(W_{PLS} S W'_{Unit}\right)$$

where CEI is a vector of the $\text{CEI}_{\text{individual}}$ values, W_{PLS} is the PLS weight matrix, and W_{Unit} is the unit weight matrix, and S is the sample correlation matrix. Using the ECSI data (Table 1), the CEI indices would be calculated as follows:

$$
\begin{bmatrix}
0.3 & 0.26 & 0.22 & 0.33 & 0.32 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.52 & 0.47 & 0.45 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.21 & 0.14 & 0.2 & 0.18 & 0.18 & 0.18 & 0.21 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.49 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.38 & 0.38 & 0.44 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.45 & 0.13 & 0.66
\end{bmatrix} \times
$$

$$
\begin{bmatrix}
1 & 0.39 & 0.26 & 0.45 & 0.43 & 0.24 & 0.28 & 0.21 & 0.48 & 0.32 & 0.45 & 0.42 & 0.37 & 0.51 & 0.47 & 0.35 & 0.38 & 0.45 & 0.38 & 0.53 & 0.42 & 0.24 & 0.02 & 0.37 \\
0.39 & 1 & 0.13 & 0.3 & 0.3 & 0.29 & 0.2 & 0.3 & 0.45 & 0.59 & 0.29 & 0.37 & 0.35 & 0.45 & 0.23 & 0.22 & 0.28 & 0.41 & 0.27 & 0.37 & 0.19 & 0.22 & 0.01 & 0.31 \\
0.26 & 0.13 & 1 & 0.41 & 0.34 & 0.22 & 0.25 & 0.12 & 0.29 & 0.11 & 0.33 & 0.2 & 0.33 & 0.24 & 0.39 & 0.28 & 0.34 & 0.22 & 0.32 & 0.29 & 0.21 & 0.25 & 0.09 & 0.28 \\
0.45 & 0.3 & 0.41 & 1 & 0.43 & 0.22 & 0.3 & 0.25 & 0.49 & 0.2 & 0.55 & 0.35 & 0.48 & 0.38 & 0.53 & 0.34 & 0.52 & 0.44 & 0.43 & 0.49 & 0.44 & 0.37 & 0.16 & 0.41 \\
0.43 & 0.3 & 0.34 & 0.43 & 1 & 0.26 & 0.36 & 0.12 & 0.46 & 0.27 & 0.51 & 0.36 & 0.54 & 0.38 & 0.39 & 0.19 & 0.3 & 0.44 & 0.39 & 0.45 & 0.34 & 0.4 & 0.06 & 0.46 \\
0.24 & 0.29 & 0.22 & 0.22 & 0.26 & 1 & 0.33 & 0.21 & 0.36 & 0.3 & 0.37 & 0.26 & 0.35 & 0.3 & 0.35 & 0.26 & 0.28 & 0.32 & 0.3 & 0.31 & 0.18 & 0.23 & 0.03 & 0.25 \\
0.28 & 0.2 & 0.25 & 0.3 & 0.36 & 0.33 & 1 & 0.11 & 0.26 & 0.17 & 0.27 & 0.31 & 0.29 & 0.33 & 0.23 & 0.14 & 0.19 & 0.37 & 0.27 & 0.28 & 0.22 & 0.25 & 0.12 & 0.3 \\
0.21 & 0.3 & 0.12 & 0.25 & 0.12 & 0.21 & 0.11 & 1 & 0.46 & 0.18 & 0.25 & 0.24 & 0.22 & 0.3 & 0.27 & 0.24 & 0.26 & 0.33 & 0.25 & 0.22 & 0.13 & 0.13 & 0.06 & 0.2 \\
0.48 & 0.45 & 0.29 & 0.49 & 0.46 & 0.36 & 0.26 & 0.46 & 1 & 0.51 & 0.52 & 0.53 & 0.52 & 0.58 & 0.56 & 0.34 & 0.5 & 0.6 & 0.57 & 0.54 & 0.38 & 0.36 & 0.02 & 0.47 \\
0.32 & 0.59 & 0.11 & 0.2 & 0.27 & 0.3 & 0.17 & 0.18 & 0.51 & 1 & 0.34 & 0.46 & 0.38 & 0.47 & 0.38 & 0.21 & 0.34 & 0.44 & 0.36 & 0.42 & 0.3 & 0.25 & -0.13 & 0.37 \\
0.45 & 0.29 & 0.33 & 0.55 & 0.51 & 0.37 & 0.27 & 0.25 & 0.52 & 0.34 & 1 & 0.52 & 0.57 & 0.49 & 0.64 & 0.4 & 0.47 & 0.46 & 0.55 & 0.59 & 0.47 & 0.31 & 0.14 & 0.48 \\
0.42 & 0.37 & 0.2 & 0.35 & 0.36 & 0.26 & 0.31 & 0.24 & 0.53 & 0.46 & 0.52 & 1 & 0.55 & 0.62 & 0.45 & 0.33 & 0.39 & 0.52 & 0.51 & 0.48 & 0.38 & 0.26 & 0.02 & 0.38 \\
0.37 & 0.35 & 0.33 & 0.48 & 0.54 & 0.35 & 0.29 & 0.22 & 0.52 & 0.38 & 0.57 & 0.55 & 1 & 0.49 & 0.51 & 0.37 & 0.47 & 0.43 & 0.44 & 0.44 & 0.39 & 0.3 & 0.07 & 0.35 \\
0.51 & 0.45 & 0.24 & 0.38 & 0.38 & 0.3 & 0.33 & 0.3 & 0.58 & 0.47 & 0.49 & 0.62 & 0.49 & 1 & 0.5 & 0.35 & 0.4 & 0.46 & 0.48 & 0.44 & 0.42 & 0.25 & 0.04 & 0.34 \\
0.47 & 0.23 & 0.39 & 0.53 & 0.39 & 0.35 & 0.23 & 0.27 & 0.56 & 0.38 & 0.64 & 0.45 & 0.51 & 0.5 & 1 & 0.47 & 0.54 & 0.49 & 0.61 & 0.64 & 0.47 & 0.34 & 0.13 & 0.43 \\
0.35 & 0.22 & 0.28 & 0.34 & 0.19 & 0.26 & 0.14 & 0.24 & 0.34 & 0.21 & 0.4 & 0.33 & 0.37 & 0.35 & 0.47 & 1 & 0.7 & 0.33 & 0.39 & 0.49 & 0.29 & 0.35 & 0.18 & 0.38 \\
0.38 & 0.28 & 0.34 & 0.52 & 0.3 & 0.28 & 0.19 & 0.26 & 0.5 & 0.34 & 0.47 & 0.39 & 0.47 & 0.4 & 0.54 & 0.7 & 1 & 0.42 & 0.51 & 0.6 & 0.36 & 0.41 & 0.08 & 0.52 \\
0.45 & 0.41 & 0.22 & 0.44 & 0.44 & 0.32 & 0.37 & 0.33 & 0.6 & 0.44 & 0.46 & 0.52 & 0.43 & 0.46 & 0.49 & 0.33 & 0.42 & 1 & 0.54 & 0.49 & 0.33 & 0.34 & 0.04 & 0.52 \\
0.38 & 0.27 & 0.32 & 0.43 & 0.39 & 0.3 & 0.27 & 0.25 & 0.57 & 0.36 & 0.55 & 0.51 & 0.44 & 0.48 & 0.61 & 0.39 & 0.51 & 0.54 & 1 & 0.59 & 0.42 & 0.35 & 0.09 & 0.49 \\
0.53 & 0.37 & 0.29 & 0.49 & 0.45 & 0.31 & 0.28 & 0.22 & 0.54 & 0.42 & 0.59 & 0.48 & 0.44 & 0.44 & 0.64 & 0.49 & 0.6 & 0.49 & 0.59 & 1 & 0.55 & 0.43 & 0.13 & 0.63 \\
0.42 & 0.19 & 0.21 & 0.44 & 0.34 & 0.18 & 0.22 & 0.13 & 0.38 & 0.3 & 0.47 & 0.38 & 0.39 & 0.42 & 0.47 & 0.29 & 0.36 & 0.33 & 0.42 & 0.55 & 1 & 0.24 & 0.12 & 0.45 \\
0.24 & 0.22 & 0.25 & 0.37 & 0.4 & 0.23 & 0.25 & 0.13 & 0.36 & 0.25 & 0.31 & 0.26 & 0.3 & 0.25 & 0.34 & 0.35 & 0.41 & 0.34 & 0.35 & 0.43 & 0.24 & 1 & 0.05 & 0.54 \\
0.02 & 0.01 & 0.09 & 0.16 & 0.06 & 0.03 & 0.12 & 0.06 & 0.02 & -0.13 & 0.14 & 0.02 & 0.07 & 0.04 & 0.13 & 0.18 & 0.08 & 0.04 & 0.09 & 0.13 & 0.12 & 0.05 & 1 & 0.1 \\
0.37 & 0.31 & 0.28 & 0.41 & 0.46 & 0.25 & 0.3 & 0.2 & 0.47 & 0.37 & 0.48 & 0.38 & 0.35 & 0.34 & 0.43 & 0.38 & 0.52 & 0.52 & 0.49 & 0.63 & 0.45 & 0.54 & 0.1 & 1
\end{bmatrix} \times
$$

$$
\begin{bmatrix}
0.29 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.29 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.29 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.29 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.29 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.48 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.48 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.48 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.19 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.54 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.54 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.48 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.48 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.48
\end{bmatrix}
=
\begin{bmatrix}
1 & 0.51 & 0.74 & 0.5 & 0.69 & 0.47 & 0.51 \\
0.51 & 1 & 0.55 & 0.36 & 0.51 & 0.26 & 0.36 \\
0.74 & 0.56 & 1 & 0.58 & 0.79 & 0.53 & 0.47 \\
0.51 & 0.36 & 0.58 & 1 & 0.6 & 0.35 & 0.5 \\
0.69 & 0.51 & 0.79 & 0.6 & 1 & 0.53 & 0.59 \\
0.46 & 0.26 & 0.53 & 0.35 & 0.52 & 1 & 0.39 \\
0.56 & 0.38 & 0.53 & 0.52 & 0.65 & 0.42 & 0.93
\end{bmatrix}
$$

The $\mathrm{CEI}_{\mathrm{individual}}$ values on the diagonal are 1.00, except for the Loyalty composite, which has a $\mathrm{CEI}_{\mathrm{individual}}$ value of 0.93, which is also the $\mathrm{CEI}_{\mathrm{minimum}}$ value.

The CEI index is applied by comparing it against a cutoff, as shown in Figure 6. Further research is required for establishing cutoffs and how the choice of cutoff affects the regression estimates but 0.95 is proposed as a conservative starting point. If there is no meaningful difference, unweighted composites are recommended, because they avoid the potential problems of PLS. Indeed, it is solid research practice to prefer simpler techniques over complex ones unless the complex technique provides a clear advantage (Kline, 2019, pp. 224–226; Wilkinson, 1999).

Small CEI values call for a choice by the researcher. In this case, researchers must interpret the weights and explain why the particular weights are sensible in the application context. If no "*a-priori* theory-based expectations of substantial differences in performance across items" (Gefen et al., 2011, p. viii) exists, equal weights should be preferred for their robustness. Indeed, as Hair et al. (2010) state "summated [sic] scales are recommended as the first choice as a remedy for measurement error where possible" (p. 172).

------ Figure 6 -------

If empirical indicator weights are used, $CEI_{individual}$ values should always be reported because this increases transparency and forces researchers to justify their choice of weights. CEI statistics are a standard part of the output in the matrixpls R package (Rönkkö, 2021) and could easily be added to other software.

FALLACY 3: USING AVE AND CR WITH PLS TO VALIDATE MEASUREMENT

A problem with PLS not related to indicator weights is the assessment of measurement quality using the Average Variance Extracted (AVE) and Composite Reliability (CR) values. Fornell and Larcker (1981) introduced the AVE and CR values to the marketing discipline as a way to evaluate confirmatory factor analysis (CFA) results. The logic of using these statistics with PLS seems to be as follows:

Premise A: PLS is a useful technique for CFA.

Premise B: AVE and CR are useful for summarizing CFA results for model assessment.

Conclusion: AVE and CR are useful for summarizing PLS results for model assessment.

Unfortunately, the conclusion is incorrect because Premise A fails. PLS does not do factor analysis. The reported "loadings" are simply correlations between indicators and composites that they form, leading to severe bias in the AVE and CR values. Yet, the use of AVE and CR continues to be advocated (e.g., Hair et al., 2020) although strong evidence against the practice has been available for a decade (Evermann & Tate, 2010), has been published in a leading research methods journal (Rönkkö & Evermann, 2013), and has been corroborated by PLS advocates' *own* research (Henseler et al., 2014; McIntosh et al., 2014)!

PLS proponents have developed two responses. The first is to deny its relevance by arguing that the studies by Evermann and coauthors (Evermann & Tate, 2010; Rönkkö & Evermann, 2013) are based on factor models, which PLS is not intended to estimate. This claim is contradicted by both the original (Jöreskog & Wold, 1982, eq. 5,7; Wold, 1982, eq. 1a-10b) and more recent PLS literature (e.g., Chin, 1998, eq. 1, 7, 9; Tenenhaus et al., 2005, pp. 163–166), which discuss factor models. Also, Hair et al. (2014) position PLS within "a class of multivariate techniques that combine aspects of factor analysis and regression" (p. xi). In research practice, PLS is used nearly exclusively for estimating factor models. To demonstrate, Google Scholar was searched for PLS-based articles published in *European Journal of Marketing* in 2020. Of the first five results, two used PLS with the AVE and CR statistics, and used the terms "loadings", "factors", and "factor loadings" (Bandara et al., 2021; Cuong et al., 2020). The other three articles (Kalra et al., 2020; Mo et al., 2020; Tarabashkina et al., 2020) explicitly used factor analysis. Against this background, claims that PLS is not used for estimating factor models are simply disingenuous.

A second and more productive approach has been the development of new model quality statistics (see Henseler, 2021). The most notable is the Heterotrait-Monotrait (HTMT) method for discriminant validity assessment (Henseler et al., 2015). Importantly, HTMT does not use PLS, but is calculated independently of any model estimates (Voorhees et al., 2016). Although abandoning PLS in favor of other methods for measurement assessment is certainly commendable, HTMT is not an ideal technique. Whereas its performance is comparable with

CFA in ideal conditions, CFA works better more generally. That Voorhees et al. (2016) report otherwise is simply due to their incorrect application of CFA (Rönkkö & Cho, 2020).

Empirical demonstration that PLS results are not useful for measurement assessment

The first row of Table 7 shows the AVE and CR values for the model shown in Panel B of Figure 2 using the corporate reputation dataset (Hair et al., 2014, Chapter 2). Following the recommended cutoffs (Hair et al., 2014), the first row would be interpreted as evidence that the *Comp*, *Like*, and *Cusa* scales are reliable and have convergent and discriminant validity.

The problem is that the model quality indices indicate a model as acceptable even when they should not: For the second row of Table 7, the model was misspecified by assigning the indicators incorrectly as shown in the first path diagram in Figure 7. This too passes the model quality heuristics with clear margins. For the third row, the Like composite was dropped and its indicators assigned to the Comp composite as shown in the second path diagram in Figure 7. Again, no problems are indicated. That is, the original analysis indicates that Comp and Like measure two different things (discriminant validity), whereas the current demonstration indicates that they measure the same thing (convergent validity). The final ten rows of Table 7 show results for models where the indicators were assigned to composites randomly (third path diagram in Figure 7). Even in these cases, the CR and AVE values *never* indicate convergent validity problems, and the AVE discriminant validity rule detects only half of the models as problematic.

------ Figure 7, Table 7 ------

Conclusions on using PLS for measurement validation

Methodological studies have used simulations shown that AVE and CR values calculated from PLS results cannot detect model misspecification. However, the same can be shown even without simulations. If these statistics are calculated from multiple different models estimated from the same data, at least some of the models are incorrect and should be identified as such, yet PLS fails to do so. PLS as a measurement validation method can thus be likened to a forecaster who always says it is going to be sunny tomorrow. It is certainly nice to hear, but ultimately useless, and you will get wet at least some of the time. In contrast, factor analysis techniques can demonstrably detect problematic models, and provide more useful input for the AVE and CR statistics.

DISCUSSION AND CONCLUSIONS

Public datasets were used to demonstrate that claims about the capabilities and advantages of PLS are either simply untrue, or at best only trivially correct. In almost every case, claims about PLS's advantages are advanced with virtually no evidence –seemingly more like marketing strategies than methodological principles. Instead of advantages, PLS comes with strong drawbacks, some of which are features of the core PLS algorithm, and no amount of ad-hoc retrofitting will remove them (Rönkkö, McIntosh, & Aguirre-Urreta, 2016).

Given that evidence of these problems has been available for years, it leads one to ask *why* there is such a big disconnect between the methodological evidence that speaks strongly against PLS, and the continued use of PLS in journals such as the *European Journal of Marketing.* PLS is an attractive method for reasons other than the quality of its results. PLS is easy to apply and will return results from virtually any dataset (Hair et al., 2011) and the model quality indices hardly ever reject the model (Rönkkö & Evermann, 2013). In a research culture that prizes publication of results over their usefulness or correctness and where there is little downside to publishing incorrect results, there are clear incentives to using PLS. As such, it falls to the reviewers and editors to challenge authors on their choice of methods (Rönkkö, McIntosh, Antonakis, et al., 2016).

What about the thousands of papers published using PLS? In the best case, the PLS weights simply do not make a difference over unit weights, and the only downside is the needlessly complicated reporting of what is essentially regression with unweighted scale scores. In other cases, PLS weights may increase the bias due to cross-loadings or inflate weak regression coefficients, producing false-positive results. Unfortunately, the weights are rarely reported, making it is difficult to assess what effect they may have had in the literature.

Using PLS to 'validate' measures may have more negative consequences that are particularly serious for newly developed scales. Scale development requires iteration because the initial scale items do not always work well (DeVellis, 2003, Chapter 5). Because these problems go undetected with PLS, the literature is contaminated with scales that are not properly validated and may not fit their intended measurement purposes. Thus, researchers are cautioned about building on prior PLS work, and encouraged to revalidate their scales with a more robust analysis before any investments in large-scale data collection.

But there are also some points of agreement between both PLS advocates and skeptics. First, weighted and unweighted composites have their uses (Rönkkö, McIntosh, Antonakis, et al., 2016, p. 2). Indeed, the first author starts his research methods course by explaining that most participants should not use SEM at all but simply use regression with unweighted composites (Rönkkö, 2020b). If used, indicator weights must serve a clear purpose aligned with research goals. Then, a suitable weighting algorithm can be chosen.

Second, PLS should be explicitly presented *only* as a composite-based technique (Henseler, 2021), that is, as an indicator weighting system, instead of using latent variable and factor analysis terminology and indices (e.g. AVE, CR). Additionally, the authors suggest dropping labels such as "structural equation modeling technique" or "second-generation multivariate technique" when discussing PLS because, regardless of their technical correctness, these labels have fundamentally misled researchers on the capabilities of the PLS technique (Rönkkö, McIntosh, Antonakis, et al., 2016; Rönkkö & Evermann, 2013). Yet, presenting PLS as "regression with weighted composites" faces two key hurdles: First, the PLS-SEM label has simply worked too well in terms of marketing the method and associated tools. Second, the more transparent labeling raises the question of the purpose of PLS weights, which the PLS literature has not answered. Ultimately, however, such a change is an essential starting point for improving empirical research in the marketing discipline.

The present article and the accompanying online supplementary material will hopefully contribute to educating researchers, reviewers, and editors on the fallacies and lesser-known facts in the use of PLS. The simple demonstrations will hopefully inspire researchers to apply them to their own datasets to advance their understanding of PLS. Hopefully, the CEI will become *de rigueur* in studies applying any composite method, especially PLS. Authors are strongly encouraged to provide more robust logic behind (and evidence for) their methodological choices, and for reviewers and editors to demand such.

APPENDIX: ONLINE SUPPLEMENTS

The article has the following supplementary material available online.

1. Alternative versions of the tables included in the article using different datasets.
2. Datasets as an excel file, including all manipulations.
3. R code that reproduces all tables and figures included in the article.
4. YouTube playlist contains screencast demonstrations and short video lectures https://youtube.com/playlist?list=PL6tc6IBlZmOWOd0OUIHkQMU3kUz1VkxY.

REFERENCES

Ali, F., Rasoolimanesh, S. M., & Cobanoglu, C. (2018). *Applying partial least squares in tourism and hospitality research*. Emerald Group Publishing.

Bandara, R., Fernando, M., & Akter, S. (2021). Managing consumer privacy concerns and defensive behaviours in the digital marketplace. *European Journal of Marketing*, *55*(1), 219–246. https://doi.org/10.1108/EJM-06-2019-0515

Becker, J.-M., Rai, A., & Rigdon, E. (2013). Predictive validity and formative measurement in structural equation modeling: Embracing practical relevance. *ICIS 2013 Proceedings*. http://aisel.aisnet.org/icis2013/proceedings/ResearchMethods/5

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Lawrence Erlbaum Associates Publishers.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–342). Sage Publications.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304

Cuong, P. H., Nguyen, O. D. Y., Ngo, L. V., & Nguyen, N. P. (2020). Not all experiential consumers are created equals: The interplay of customer equity drivers on brand loyalty. *European Journal of Marketing*, *54*(9), 2257–2286. https://doi.org/10.1108/EJM-04-2018-0228

DeVellis, R. F. (2003). *Scale development theory and applications*. Sage.

Dijkstra, T. K. (2009, September 4). PLS for path diagrams revisited, and extended. *Proceedings of the 6th International Conference on Partial Least Squares*. 6th International Conference on Partial Least Squares, Beijing.

Evermann, J., & Rönkkö, M. (2021). Recent developments in PLS. *Communications of the Association for Information Systems*, *forthcoming*.

Evermann, J., & Tate, M. (2010). Testing models or fitting models? Identifying model misspecification in PLS. *ICIS 2010 Proceedings*. http://aisel.aisnet.org/icis2010_submissions/21

Fornell, C. (1987). A second generation of multivariate analysis: Classification of methods and implications for marketing research. *Review of Marketing*, *87*, 407–450.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, *19*(4), 440–452.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Gefen, D., Rigdon, E. E., & Straub, D. W. (2011). An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly*, *35*(2), iii–xiv.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data. *MIS Quarterly*, *36*(3), 981–1001.

Goodhue, D. L., Lewis, W., & Thompson, R. (2015). PLS pluses and minuses in path estimation accuracy. *AMCIS 2015 Proceedings*. http://aisel.aisnet.org/amcis2015/ISPhil/GeneralPresentations/3

Graefe, A. (2015). Improving forecasts using equally weighted predictors. *Journal of Business Research*, *68*(8), 1792–1799. https://doi.org/10.1016/j.jbusres.2015.03.038

Grice, J. W. (2001). A comparison of factor scores under conditions of factor obliquity. *Psychological Methods*, *6*(1), 67–83. https://doi.org/10.1037//1082-989X.6.1.67

Hair, J. F. (2010). *Multivariate data analysis: A global perspective.* (7th ed.). Pearson Education.

Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research*, *109*, 101–110. https://doi.org/10.1016/j.jbusres.2019.11.069

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equations modeling (PLS-SEM)*. SAGE Publications.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science*, 1–17. https://doi.org/10.1007/s11747-017-0517-x

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory & Practice*, *19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Henseler, J. (2021). *Composite-based structural equation modeling: Analyzing latent and emergent variables*. The Guilford Press.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Hair, J. F., Hult, G. T. M., & Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, *17*(2), 182–209. https://doi.org/10.1177/1094428114526928

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, *28*(2), 565–580. https://doi.org/10.1007/s00180-012-0317-1

Jöreskog, Karl. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction*. North-Holland.

Kalra, A., Agnihotri, R., Singh, R., Puri, S., & Kumar, N. (2020). Assessing the drivers and outcomes of behavioral self-leadership. *European Journal of Marketing*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/EJM-11-2018-0769

Kline, R. B. (2019). *Becoming a behavioral science researcher: A guide to producing research that matters* (Second edition). The Guilford Press.

Lee, N., Cadogan, J. W., & Chamberlain, L. (2013). The MIMIC model and formative variables: Problems and solutions. *AMS Review*, *3*(1), 3–17. https://doi.org/10.1007/s13162-013-0033-1

McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on partial least squares path modeling. *Organizational Research Methods*, *17*(2), 210–251. https://doi.org/10.1177/1094428114529165

Mo, C. (Jacky), Yu, T., & de Ruyter, K. (2020). Don't you (forget about me): The impact of out-of-the-channel-loop perceptions in distribution channels. *European Journal of Marketing*, *54*(4), 761–790. https://doi.org/10.1108/EJM-05-2018-0324

Nunnally, J. (1978). *Psychometric theory*. McGraw-Hill.

Ravand, H., & Baghaei, P. (2016). Partial least squares structural equation modeling with R. *Review of Practical Assessment, Research, & Evaluation*, *21*(11).

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30–45. https://doi.org/10.1037/met0000220

Rigdon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal*, *34*(6), 598–605. https://doi.org/10.1016/j.emj.2016.05.006

Rönkkö, M. (2014). The effects of chance correlations on partial least squares path modeling. *Organizational Research Methods*, *17*(2), 164–181. https://doi.org/10.1177/1094428114525667.

Rönkkö, M. (2020a). *Introduction to matrixpls*. https://cran.r-project.org/web/packages/matrixpls/vignettes/matrixpls-intro.pdf

Rönkkö, M. (2020b, December 1). *Back to basics approach*. https://www.youtube.com/watch?v=ROiClgEz2No

Rönkkö, M. (2021). *matrixpls: Matrix-based partial least squares estimation* (1.0.12) [Computer software]. https://github.com/mronkko/matrixpls

Rönkkö, M., & Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organizational Research Methods*. https://doi.org/10.1177/1094428120968614

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*(3), 425–448. https://doi.org/10.1177/1094428112474693

Rönkkö, M., Evermann, J., & Aguirre-Urreta, M. I. (2016). Estimating formative measurement models in IS research: Analysis of the past and recommendations for the future. *Unpublished Working Paper*. http://urn.fi/URN:NBN:fi:aalto-201605031907

Rönkkö, M., McIntosh, C. N., & Aguirre-Urreta, M. I. (2016). Improvements to PLSc: Remaining problems and simple solutions. *Unpublished Working Paper*. http://urn.fi/URN:NBN:fi:aalto-201603051463

Rönkkö, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences*, *87*, 76–84. https://doi.org/10.1016/j.paid.2015.07.019

Rönkkö, M., McIntosh, C. N., Antonakis, J., & Edwards, J. R. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management*, *47–48*, 9–27. https://doi.org/10.1016/j.jom.2016.05.002

Rönkkö, M., & Ylitalo, J. (2010). Construct validity in partial least squares path modeling. *ICIS 2010 Proceedings*. http://aisel.aisnet.org/icis2010_submissions/155

SmartPLS. (2020). *Technology Acceptance Model Project*. https://www.smartpls.com/documentation/sample-projects/tam

Tarabashkina, L., Quester, P. G., & Tarabashkina, O. (2020). How much firms "give" to CSR vs how much they "gain" from it: Inequity perceptions and their implications for CSR authenticity. *European Journal of Marketing*, *54*(8), 1987–2012. https://doi.org/10.1108/EJM-11-2018-0772

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, *48*(1), 159–205. https://doi.org/10.1016/j.csda.2004.03.005

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, *44*(1), 119–134. https://doi.org/10.1007/s11747-015-0455-4

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, *40*(5), 663–705. https://doi.org/10.3102/00346543040005663

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & S. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 1–54). North-Holland.

TABLES

Table 1 Correlations between the PLS composites and unit weighted composites using the ECSI data and model

		PLS Mode A composites							Unit weight composites							
		1	2	3	4	5	6	7	8	9	10	11	12	13	15	16
1	Image	1														
2	Expectation	0.505	1													
3	Quality	0.749	0.557	1												
4	Value	0.508	0.361	0.586	1											
5	Satisfaction	0.693	0.510	0.795	0.606	1										
6	Complaints	0.475	0.257	0.532	0.355	0.526	1									
7	Loyalty	0.564	0.380	0.538	0.530	0.656	0.418	1								
8	Image	**0.997**	0.507	0.744	0.510	0.685	0.463	0.557	1							
9	Expectation	0.506	**0.999**	0.557	0.361	0.510	0.257	0.380	0.508	1						
10	Quality	0.744	0.554	**0.999**	0.578	0.788	0.528	0.533	0.739	0.553	1					
11	Value	0.501	0.359	0.579	**0.999**	0.599	0.351	0.524	0.503	0.359	0.572	1				
12	Satisfaction	0.690	0.512	0.794	0.600	**0.999**	0.519	0.652	0.683	0.513	0.788	0.593	1			
13	Complaints	0.475	0.257	0.532	0.355	0.526	**1.000**	0.418	0.463	0.257	0.528	0.351	0.519	1		
14	Loyalty	0.513	0.355	0.473	0.499	0.585	0.386	**0.932**	0.507	0.356	0.466	0.497	0.580	0.386	1	
16	Loyalty, excl item 2	0.554	0.370	0.528	0.516	0.637	0.390	**0.986**	0.548	0.370	0.525	0.510	0.632	0.390	0.879	1

Note. Correlations between a PLS composite and corresponding unweighted composite are bolded.

Table 2 Comparison of R^2 between PLS Mode A, PLS Mode B, and canonical correlation weights using the ECSI dataset

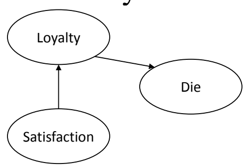
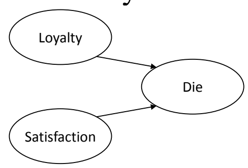
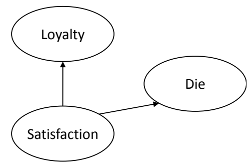
	PLS Mode A	PLS Mode B	Canonical correlation weights
Composite weights			
imag1	0.263	0.138	0.521
imag2	0.226	0.204	-0.118
imag3	0.229	0.123	-0.266
imag4	0.342	0.383	-0.251
imag5	0.366	0.520	-0.850
cusa1	0.371	0.347	-0.351
cusa2	0.366	0.146	-0.107
cusa3	0.462	0.676	-0.703
cusco	1	1	1
cusl1	0.452	0.208	-0.244
cusl2	0.133	0.105	-0.092
cusl3	0.658	0.856	-0.833
Regression of Loyalty			
Image	0.206	0.183	0.201
Satisfaction	0.481	0.524	0.560
Complaints	0.066	0.074	-0.086
R^2	0.462	0.499	0.514

Table 3 Effects of chance correlations on PLS weights and composite correlations

	Equal correlations		Unequal correlations	
	PLS Mode A	Equal weights	PLS Mode A	Equal weights
Indicator weights				
a1, b1	0.410	0.410	0.506	0.410
a2, b2	0.410	0.410	0.324	0.410
a3, b3	0.410	0.410	0.395	0.410
Composite correlation	0.223	0.223	0.236	0.223

Note: Models based on Figure 3. Weights of a and b indicators are symmetric.

Table 4 Weights and composite correlations for three different PLS analyses and unit weights using the ECSI dataset

	PLS analysis 1	PLS analysis 2	PLS analysis 3	Unit weights
				
Composite correlations				
Loyalty - Satisfaction	**0.650**	0.022	**0.659**	0.580
Die - Satisfaction	0.012	**-0.160**	**-0.044**	-0.009
Die - Loyalty	**-0.088**	**-0.109**	0.023	-0.048
Indicator weights				
cusa1	0.370	1.165	0.437	0.400
cusa2	0.367	-0.814	0.318	0.400
cusa3	0.462	0.015	0.446	0.400
cusl1	0.505	0.973	0.453	0.478
cusl2	0.136	0.551	0.104	0.478
cusl3	0.609	-0.385	0.663	0.478
die1	0.972	0.340	-0.302	0.407
die2	0.255	0.428	0.826	0.407
die3	-0.275	0.450	0.465	0.407

Note. Adjacent PLS composites are bolded

Table 5 Factor analysis of the corporate reputation dataset

Indicator	Factors		
	Factor 1: Like	Factor 2: Cusl	Factor 3: Comp
comp1	**0.412**	0.083	**0.302**
comp2	-0.017	0.033	**0.777**
comp3	0.017	-0.027	**0.833**
like1	**0.796**	0.015	0.031
like2	**0.846**	-0.007	-0.071
like3	**0.655**	0.033	0.096
cusl1	**0.210**	**0.548**	0.043
cusl2	-0.025	**0.983**	-0.023
cusl3	-0.003	**0.713**	0.038

Note. Principal axis factoring with oblimin rotation. Loadings exceeding 0.1 in absolute value are bolded.

Table 6 Regressions with PLS and unit weighted composites using the corporate reputation dataset

	Original data		Data with artificially generated cross-loading	
	PLS Mode A	Unit weights	PLS Mode A	Unit weights
Regression estimates				
Comp → Cusa	0.152	0.122	0.058	0.045
Comp → Cusl	0.016	0.011	-0.095	-0.074
Like → Cusa	0.433	0.452	0.578	0.570
Like → Cusl	0.331	0.333	0.599	0.535
Cusa → Cusl	0.509	0.511	0.364	0.401
Weights				
comp1	0.539	0.401	0.539	0.401
comp2	0.343	0.401	0.343	0.401
comp3	0.323	0.401	0.323	0.401
like1	0.419	0.386	0.348	0.382
like2	0.378	0.386	0.489	0.382
like3	0.360	0.386	0.300	0.382
cusa	1.000	1.000	1.000	1.000
cusl1	0.368	0.385	0.373	0.385
cusl2	0.418	0.385	0.416	0.385
cusl3	0.366	0.385	0.363	0.385

Table 7 Comparing the AVE and CR statistics for the original model and twelve misspecified models using the corporate reputation dataset

	CR			AVE			AVE – largest squared correlation		
	Comp	Like	Cusl	Comp	Like	Cusl	Comp	Like	Cusl
Original	0.864	0.898	0.900	0.680	0.746	0.751	0.263	0.329	0.273
Misspecified 1	0.793	0.852	0.899	0.562	0.661	0.749	0.044	0.143	0.249
Misspecified 2	0.892		0.900	0.581		0.751	0.581		0.273
Random 1	0.869	0.811	0.831	0.689	0.589	0.621	0.104	0.009	0.037
Random 2	0.831	0.804	0.868	0.623	0.583	0.687	0.084	**-0.021**	0.083
Random 3	0.816	0.864	0.785	0.600	0.680	0.550	**-0.026**	0.111	**-0.075**
Random 4	0.843	0.822	0.900	0.642	0.609	0.750	0.080	0.047	0.273
Random 5	0.874	0.838	0.806	0.698	0.636	0.581	0.119	0.049	-0.006
Random 6	0.818	0.839	0.840	0.601	0.634	0.637	**-0.024**	**-0.015**	**-0.013**
Random 7	0.806	0.820	0.875	0.582	0.603	0.699	0.058	**-0.048**	0.048
Random 8	0.806	0.848	0.852	0.584	0.651	0.658	0.016	0.082	0.093
Random 9	0.797	0.809	0.898	0.571	0.587	0.746	**-0.044**	**-0.028**	0.213
Random 10	0.817	0.847	0.854	0.601	0.649	0.662	0.000	0.047	0.086

Note. Bolded values would be considered as problematic.

FIGURES

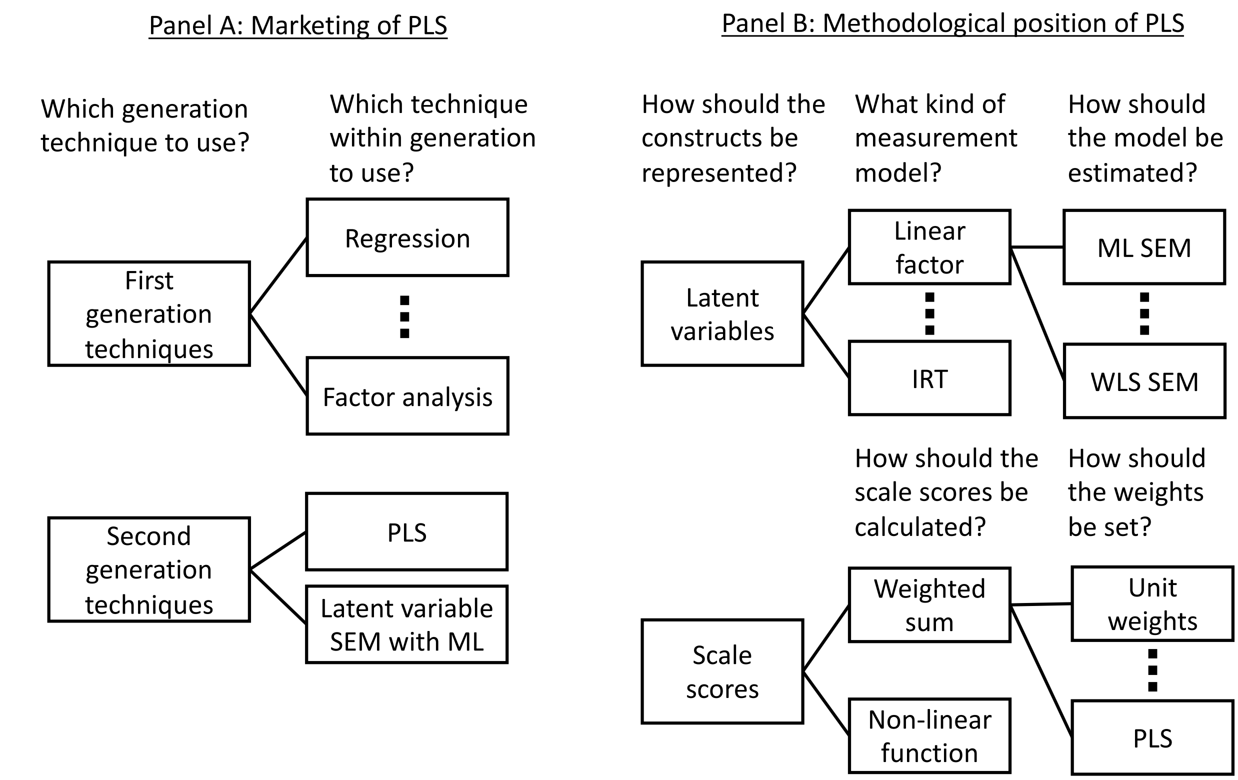


Figure 1 Comparison of how PLS is marketed and how it positions methodologically

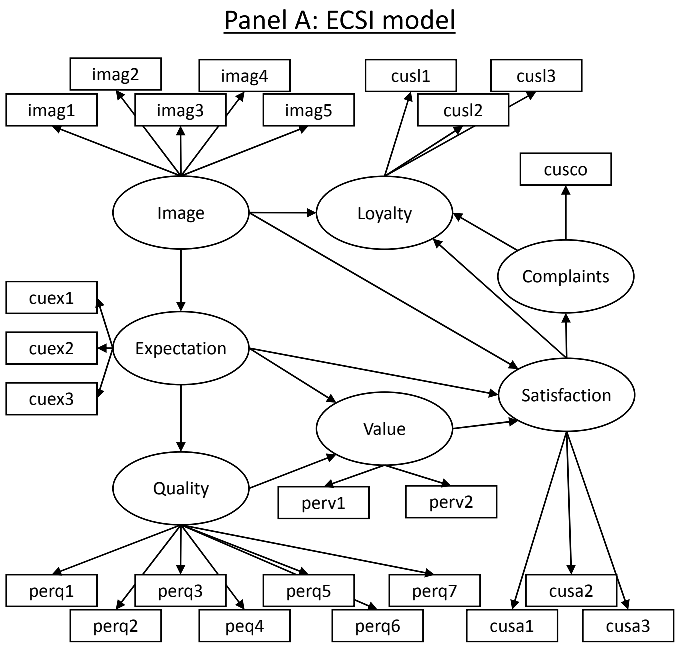


Figure 2 Path diagrams of the example model

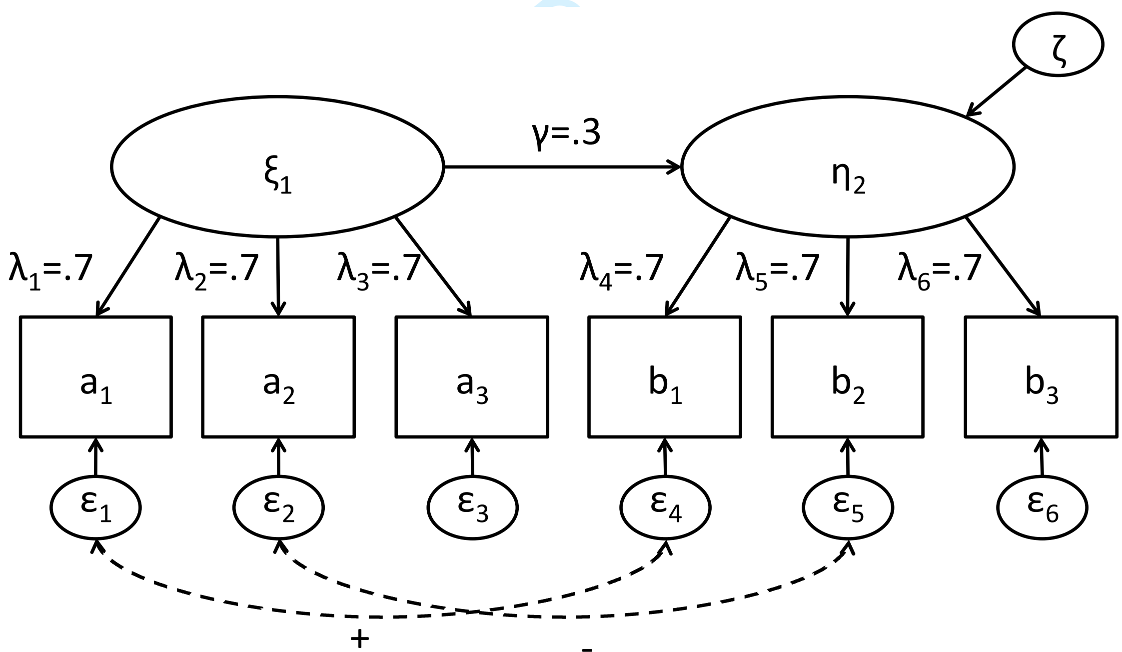


Figure 3 Example of chance correlations

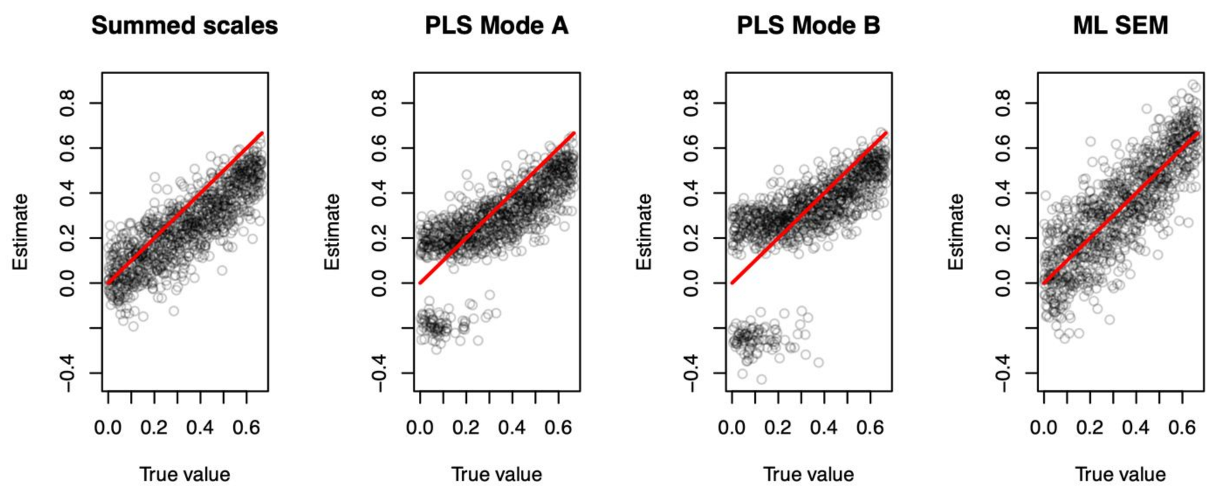


Figure 4 Comparison of regression with unweighted composites, PLS composites, and latent variable SEM using 1000 replications of simulated data.

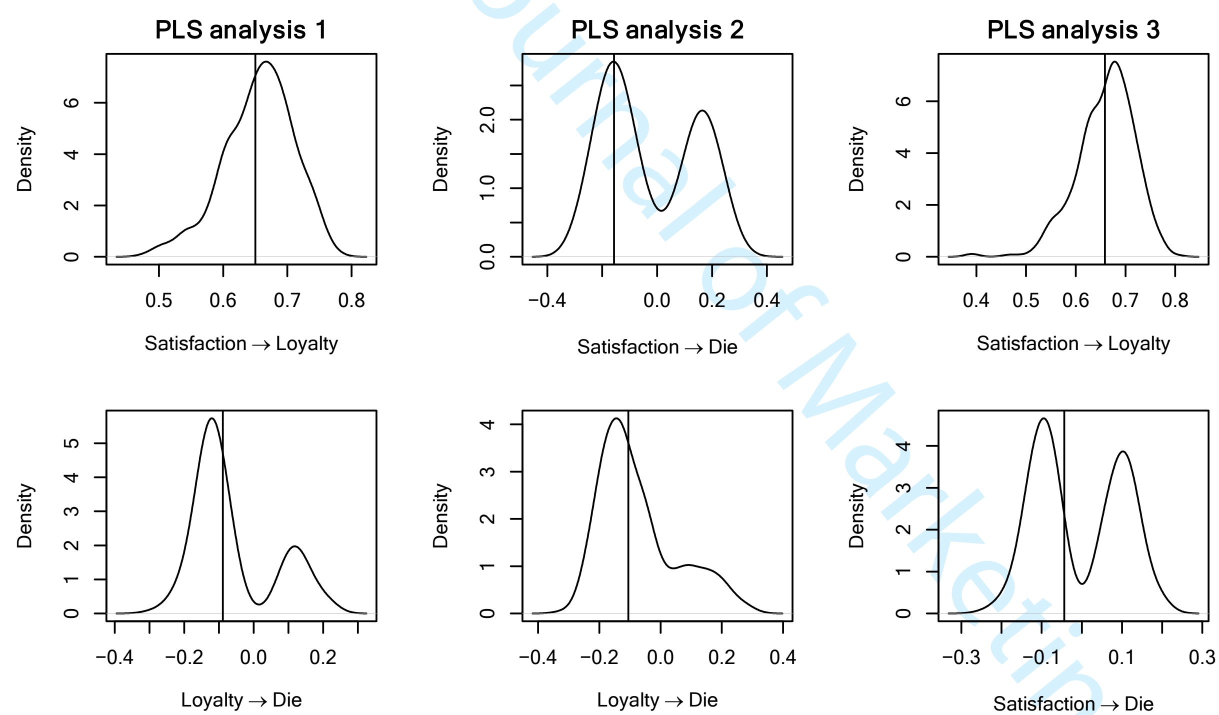


Figure 5 Bootstrap distributions of the estimates for the three example PLS analyses. The vertical line marks the original estimate.

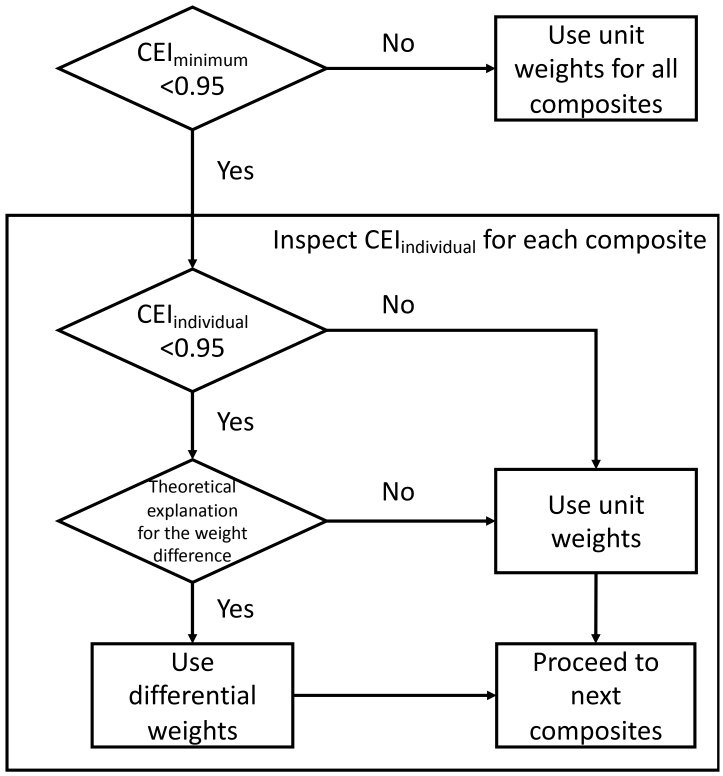


Figure 6 Guidelines for choosing between unit weights and differential weights based on CEI statistics

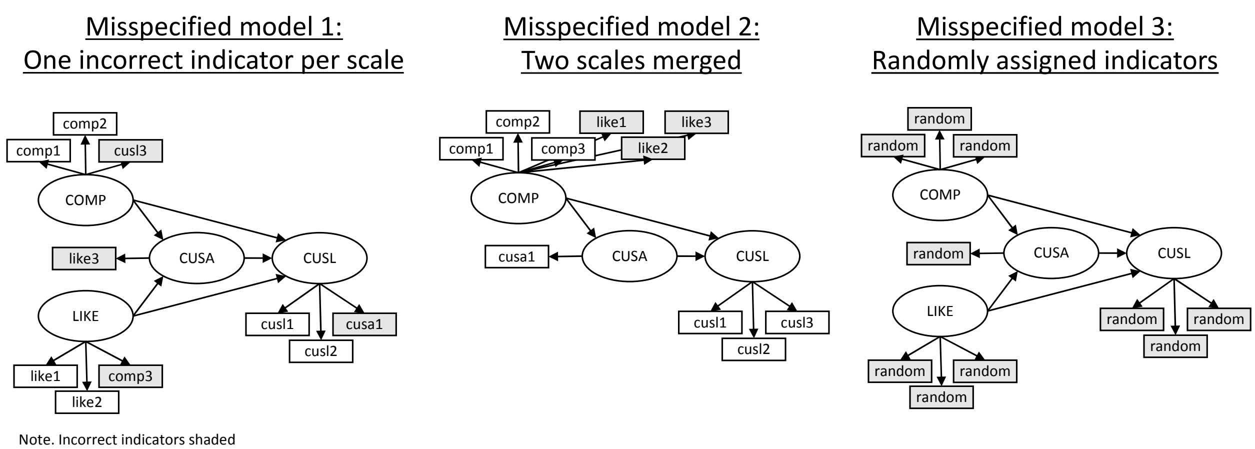


Figure 7 Three misspecified corporate reputation models

21 June 2021

Dear Dr. Voss,

Thank you for the opportunity to improve our manuscript and for the constructive and helpful comments by the guest editor. We have edited our manuscript in response to those comments and we hope that the improved version will meet with your approval for publication in the special issue of EJM. In the table below we respond to specific review comments.

Kind regards,

The authors

Guest editor comment	Our response.
In the previous, I questioned the cut-off criteria for the CEI. I appreciate the changes that were made. The section is improved. However, your reply, in particular, the stock market average price calculations made clear that there must be a distinction between reflective and formative scales. I request you consider that the suggested cut-off for a reflective scale may not be the same as for a formative scale.	This issue is a bit more nuanced than what we can explain in the article because of length limitations. As such, please allow us to explain it here, in the hope that this explanation will reassure you that the discussion in the main paper is adequate. First, items in a valid and reliable reflective scale should always be highly correlated. If they are not correlated, this is either a reliability or a validity problem and should be caught by a factor analysis that should always be run prior to calculating composites from reflective scales. With highly correlated items like this, it is difficult to see how the CEI statistic would ever indicate a difference. (Unless one composite is specifically designed to be very different, by for example using a mix of positive and negative weights.) Because of this fact, the existing literature on scale score construction does not even recommend comparing different composite scores, but just to default to unit weights. That being said, to educate the readers on the (lack of) potential advantages of indicator weighs, we think that asking them to calculate CEI and seeing it for themselves is much more effective than asking them to read reviews by Cohen or someone else.

	With that in mind, we think that the CEI index is more useful for formative scales, because at least in principle with these scales weights can make a difference. While as stated in the article, they often do not, they could and in that case it is essential that researchers provide a theoretical justification for choosing one set of weights over another. Unfortunately, users of formative scales almost never justify their weighting schemes, so in this sense CEI should be a major help. In all, we could view CEI and the proposed cutoffs applicable *only* to formative indicators. (Because reflective indicators should always be aggregated as a composite using unit weights.) However, the reason we have not discussed this in the paper is that we do not want to focus on formative indicators because a) it would complicate the paper and b) we see value in readers calculating CEIs themselves regardless of what data they use to learn when indicator weights make a difference and when they do not, even if we know *a priori* that weights should not make a difference for well-behaving reflective scales. Second, we do not ourselves see why it would matter that kind of measurement assumptions we make about the scales. If two composites are nearly perfectly correlated, they are equivalent empirically, which the CEI measures. As an analogy, if we wanted to compare different people to see who has the most money, for answering that question it does not matter if that money originated from hard work, inheritance, or criminal activity. Because of these considerations, we decided not to adjust our recommended cutoff. Of course, further research can and probably should do so.

	We would be of course willing to add some discussion of this if you felt it was absolutely necessary, perhaps as a footnote om p14. However, as already stated we are absolutely on the word count limit already, which has been achieved only through a line by line reworking. We feel that taking content out of the paper would not be justified by adding this discussion, which is to some extent practically marginal, and may even confuse readers if we do not extend it in enough detail. So, in order to add this, we would probably need between 50 (for a footnote) to a couple of hundred extra words (for an in-text discussion) over the count.
There is a number of different treatments in the manuscript concerning the page numbers for quoted text: Introduction, para 1 line 4: contained within cited reference quote	Thanks for these detailed suggestions, however, the journal style guide (https://www.emeraldgrouppublishing.com/journal/ejm?distinct_id=17734cb504a45d-0ae92887505eb2-48193201-384000-17734cb504bc44&_ga=2.182797602.524475136.1623825722-380842808.1611498476#author-guidelines) does not give any specific guidance on how to use citations with quoted materials. The guidelines state that "All references in your manuscript must be formatted using one of the recognised Harvard styles" and further "Want to use a different Harvard style? That's fine, our typesetters will make any necessary changes to your manuscript if it is accepted" As such, we used the American Psychological Association style, 7th edition. More specifically, we follow the standard set in Section 8.26 of APA Manual (American Psychological Association, 2019), because this format is what the author team write most of their work in. We use both parenthetical and narrative citations. When using parenthetical citations, the citation follows in full directly after the quote as in: "quote" (author, year, p. page).

	In narrative citations, the author and the year are always presented earlier in the sentence and the page number appears in quotes after the quote, as in: Author (year) writes “quote” (p. page). We found one quote where this standard was not followed because the page number was presented before the quoted material. This has been corrected. Additionally, we corrected a couple of misplaced periods. We can convert the article to use something other than the APA format (for example, by placing the page numbers within the same parens for narrative quotes), but would need more specific guidance on how exactly quotes should be marked in the case of both narrative and parenthetical citations. That said, it seems according to the EJM website quote above, that the typesetters will deal with this if necessary.
Introduction, para 2 line 13: separate parens at the end of the quote before the period.	We have not made any changes here, since (as above) we are not sure where the problem lies.
fallacy 1, para 1, line 3: separate parens at the end of the quote before the period.... there is a period at the end of the quote as well.	We have removed an extra period inside the quote.
Fallacy 2: para 1, line 10: separate parens at the end of the quote.	We have not made any changes here, since (as above) we are not sure where the problem lies.
Fallacy 2, para 1, line 11: separate parens outside of the sentence punctuation.	We have not made any changes here, since (as above) we are not sure where the problem lies.
Fallacy 2, para 2, line 3 and 11: contained within cited reference quote.	The page number has been moved to the end of the quote.
Chance correlations, para 2, line 6: separate parens outside of the sentence punctuation.	We have not made any changes here, since (as above) we are not sure where the problem lies.

Empirical demonstration of cross-loadings: para 2, line 10: contained within cited reference quote.	We have not made any changes here, since (as above) we are not sure where the problem lies.

ONLINE SUPPLEMENT 1

In the article, we used three different empirical data sets to demonstrate problems in the PLS technique. For each demonstration, we only presented one set of results in the article. This online supplement presents alternative versions of Table 1 using different datasets.

The first data set is the ECSI data set that is provided as an example in many PLS software packages and is used, in different versions, in various articles. We use the version adapted to mobile phone markets, from Tenenhaus et al. (2005). The data consist of 250 observations on 24 variables. The second data set is from the "corporate reputation" example in Hair et al. (2014, Chapter 2) and consists of 336 observations on 10 variables. The third data set is the TAM data set from the SmartPLS website (SmartPLS, 2020) and consists of 1190 observations on 23 variables.

REFERENCES

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equations modeling (PLS-SEM)*. SAGE Publications.

SmartPLS. (2020). *Technology Acceptance Model Project*. https://www.smartpls.com/documentation/sample-projects/tam

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, *48*(1), 159–205. https://doi.org/10.1016/j.csda.2004.03.005

Table S1 Correlations between the PLS composites and unit weighted composites using the ECSI data and model (Table 1 in the article)

		PLS Mode A composites							Unit weight composites							
		1	2	3	4	5	6	7	8	9	10	11	12	13	15	16
1	Image	1														
2	Expectation	0.505	1													
3	Quality	0.749	0.557	1												
4	Value	0.508	0.361	0.586	1											
5	Satisfaction	0.693	0.510	0.795	0.606	1										
6	Complaints	0.475	0.257	0.532	0.355	0.526	1									
7	Loyalty	0.564	0.380	0.538	0.530	0.656	0.418	1								
8	Image	**0.997**	0.507	0.744	0.510	0.685	0.463	0.557	1							
9	Expectation	0.506	0.**999**	0.557	0.361	0.510	0.257	0.380	0.508	1						
10	Quality	0.744	0.554	**0.999**	0.578	0.788	0.528	0.533	0.739	0.553	1					
11	Value	0.501	0.359	0.579	**0.999**	0.599	0.351	0.524	0.503	0.359	0.572	1				
12	Satisfaction	0.690	0.512	0.794	0.600	**0.999**	0.519	0.652	0.683	0.513	0.788	0.593	1			
13	Complaints	0.475	0.257	0.532	0.355	0.526	**1.000**	0.418	0.463	0.257	0.528	0.351	0.519	1		
14	Loyalty	0.513	0.355	0.473	0.499	0.585	0.386	**0.932**	0.507	0.356	0.466	0.497	0.580	0.386	1	
16	Loyalty, excl item 2	0.554	0.370	0.528	0.516	0.637	0.390	**0.986**	0.548	0.370	0.525	0.510	0.632	0.390	0.879	1

Note. Correlations between a PLS composite and corresponding unit weight composite are bolded.

Table S2 Correlations between the PLS composites and unit weighted composites using the corporate reputation data and model

		PLS Mode A composites				Unit weight composites			
		1	2	3	4	5	6	7	8
1	Comp	1							
2	Like	0.646	1						
3	Cusa	0.432	0.531	1					
4	Cusl	0.450	0.612	0.692	1				
5	Comp	**0.992**	0.629	0.406	0.426	1			
6	Like	0.644	**1.000**	0.529	0.611	0.627	1		
7	Cusa	0.432	0.531	**1.000**	0.692	0.406	0.529	1	
8	Cusl	0.451	0.611	0.691	**1.000**	0.427	0.610	0.691	1

Note. Correlations between a PLS composite and corresponding unit weight composite are bolded.

Table S3 Correlations between the PLS composites and unit weighted composites using the TAM data and model

		PLS Mode A composites					Unit weight composites				
		1	2	3	4	5	6	7	8	9	10
1	USEF	1									
2	EOU	0.443	1								
3	BI	0.453	0.547	1							
4	ATT	0.379	0.369	0.316	1						
5	USE	0.226	0.326	0.213	0.314	1					
6	USEF	**1.000**	0.444	0.453	0.379	0.225	1				
7	EOU	0.443	**1.000**	0.547	0.369	0.325	0.444	1			
8	BI	0.455	0.548	**1.000**	0.318	0.213	0.455	0.548	1		
9	ATT	0.379	0.369	0.316	**1.000**	0.314	0.379	0.369	0.318	1	
10	USE	0.225	0.326	0.211	0.315	**0.999**	0.225	0.324	0.211	0.315	1

Note. Correlations between a PLS composite and corresponding unit weight composite are bolded.