The shortcomings of equal weights estimation and the composite equivalence index in PLS-SEM

Joseph F. Hair
Department of Marketing, University of South Alabama, Mobile, Alabama, USA

Pratyush N. Sharma
Department of Information Systems, Statistics and Management Science, Culverhouse College of Business, The University of Alabama, Tuscaloosa, Alabama, USA

Marko Sarstedt
Institute for Marketing, Ludwig-Maximilians-Universität München, München, Germany and Faculy of Economics and Business Administration, Babeș-Bolyai University, Cluj-Napoca, Romania

Christian M. Ringle
Department of Management Sciences and Technology, Hamburg University of Technology, Hamburg, Germany, and

Benjamin D. Liengaard
Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark

Abstract

Purpose – The purpose of this paper is to assess the appropriateness of equal weights estimation (sumscores) and the application of the composite equivalence index (CEI) *vis-à-vis* differentiated indicator weights produced by partial least squares structural equation modeling (PLS-SEM).

Design/methodology/approach – The authors rely on prior literature as well as empirical illustrations and a simulation study to assess the efficacy of equal weights estimation and the CEI.

Findings – The results show that the CEI lacks discriminatory power, and its use can lead to major differences in structural model estimates, conceals measurement model issues and almost always leads to inferior out-of-sample predictive accuracy compared to differentiated weights produced by PLS-SEM.

Research limitations/implications – In light of its manifold conceptual and empirical limitations, the authors advise against the use of the CEI. Its adoption and the routine use of equal weights estimation could adversely affect the validity of measurement and structural model results and understate structural model predictive accuracy. Although this study shows that the CEI is an unsuitable metric to decide between equal weights and differentiated weights, it does not propose another means for such a comparison.

Practical implications – The results suggest that researchers and practitioners should prefer differentiated indicator weights such as those produced by PLS-SEM over equal weights.

Originality/value – To the best of the authors' knowledge, this study is the first to provide a comprehensive assessment of the CEI's usefulness. The results provide guidance for researchers considering using equal indicator weights instead of PLS-SEM-based weighted indicators.

Keywords Composite equivalence index, CEI, Equal weights, Measurement model, Partial least squares, PLS-SEM, Structural equation modeling

Paper type Technical paper

1. Introduction
Structural equation modeling using partial least squares (PLS-SEM) has traditionally relied on differentiated indicator weights for estimating models that comprise structural relationships between constructs as statistical representations of unobservable concepts (Wold, 1982; Lohmöller, 1989; Hair *et al.*, 2022). By assigning different weights to a construct's indicators, researchers not only account for differences in their relevance (Rigdon, 2012) but also correct for measurement error inherent in the indicators (Henseler *et al.*, 2014). The recent surge in methodological developments has resulted in several empirical metrics and recommendations worthy of scholarly discussion regarding the appropriateness of their adoption in the PLS-SEM framework. The purpose of this paper is to assess one such recent recommendation – the proposed adoption of the composite equivalence index (CEI) and the preferred use of equal weights over differentiated weights to compute composite scores for estimating the structural model relationships (Rönkkö *et al.*, 2023) [1]. We demonstrate why the issues discussed in this paper render the CEI unsuitable for adoption within the PLS-SEM framework under most practical conditions and emphasize the benefits of preferring differentiated weights over equal weights.

As a background, the CEI represents a simple index for assessing the correlation between the composite scores produced by two scoring methods:

(1) equal weights (a.k.a. sumscores, created by equally weighting all indicators of a construct); and

(2) differentiated weights obtained by applying PLS-SEM (i.e. weighted scores created by assigning differential weights to the indicators of a construct) [2].

A CEI value close to unity indicates a high degree of similarity between the composite scores generated by equal weights and those computed on the grounds of differentiated weights. The proposed CEI metric recommends using equal weights over the differentiated PLS-SEM-based weights when its value is larger than 0.95 (i.e. when there is a high correlation between the sumscores and weighted PLS-SEM scores; Rönkkö *et al.*, 2023). In this case, the proposed recommendation is for researchers to compute sumscores and use them as input for estimating the partial regressions in the structural model instead of PLS-SEM-based weights. When the CEI value is below 0.95, the proposed recommendation is for researchers to use the PLS-SEM weights, but only if researchers can offer an *a priori* theory-based justification for the expected difference in indicator weights.

The debate broached by the CEI regarding the use of equal versus differentiated weights is far from new and has been repeatedly raised in prior psychometric research (e.g. McNeish and Wolf, 2020; McNeish, 2023, Widaman and Revelle, 2023a, 2023b) – albeit only recently in the PLS-SEM context. Researchers endorsing the use of equal weights frequently emphasize that these scores are typically highly correlated with differentially weighted scores (which translates into high CEI values), potentially producing negligible differences when used in the follow-up analyses (Bobko *et al.*, 2007; Ree *et al.*, 1998; Wainer, 1976). This is summarized by Kline (2005, p. 105), who notes that "the correlation between weighted and unit-weighted [summed] test scores is almost 1.0. Thus, the take-home message is pretty simple – don't bother to differentially weight items. It is not worth the effort."

This take-home message, which is also the main logic behind the proposal of the CEI, is too simplistic. As researchers have developed more understanding about this issue in recent years, the downsides of using sumscores have become apparent. For example, in their article, "Thinking twice about sum scores," McNeish and Wolf (2020) noted several empirical and conceptual shortcomings of equal weights including, but not limited to:

- equal weights produce unrealistic expectations about the population model by enforcing unnatural constraints on the empirical model, which prove particularly problematic in the context of measurement invariance assessment;
- they hinder rigorous and accurate psychometric assessments by ignoring measurement theory in its entirety;
- they adversely affect construct validity and reliability;
- they can result in vastly different conclusions due to inaccurate coefficient estimation; and
- because virtually all the psychometric scales used in management and social sciences research have been validated under the assumption of differentiated weights, using equal weights when applying these scales is categorically inappropriate because doing so enforces a model different from the one validated.

Hence, assuming equal weights estimation as a quasi-default – as practically implied by the CEI – reintroduces various problems that have long been overcome through the implementation of estimators that do not impose artificial constraints on the model.

More importantly, McNeish (2023) shows that analyses based on equal versus differentiated weights can cause large variability in conclusions, even when the corresponding scores are very highly correlated – higher than the suggested CEI threshold of 0.95. Specifically, differentiated weights produce scores that have higher correlations with true scores, higher sensitivity and higher reliability even at correlations as high as 0.98. These findings tie in with extensive literature showcasing the inadequacy of equal weights compared to differentiated weights (e.g. Lastovicka and Thamodaran, 1991; Andersson and Yang-Wallentin, 2020; Rigdon *et al.*, 2019). In other words, a high correlation between equally weighted scores and differentially weighted scores, as reflected in high CEI values, provides no guarantee of comparable results. The mere fact that sumscores are highly correlated with differentiated scores does not mean they are interchangeable and that there is no penalty to pay when using sumscores (e.g. Estabrook and Neale, 2013; Hair *et al.*, 2017b; Murray *et al.*, 2016). Not surprisingly, researchers have noted that sumscores are too imprecise to be used in rigorous empirical research in most conditions – and that their use

should be restricted to very specific conditions (e.g. McNeish and Wolf, 2020; McNeish, 2023).

In what follows, we extend these arguments and demonstrate how the application of the CEI leads to substantial drawbacks since the index suffers from multiple conceptual and empirical shortcomings. We first show that the CEI lacks discriminatory power in that the suggested thresholds do not differentiate between different outcomes, but univocally favor sumscores, which, as we will show, has adverse consequences for findings' practical relevance. We then show that researchers applying the CEI are likely to overlook the problems caused by the unreliable indicators, whose inclusion can have profound effects on the structural model results. Finally, extending on current discussions in the psychometrics literature (e.g. McNeish and Wolf, 2020; McNeish, 2023; Widaman and Revelle, 2023b), we show by means of a simulation study and an empirical example that following the recommendations implied by the CEI will generally lead to models with inferior out-of-sample predictive accuracy – an important goal in PLS-SEM analyses (Danks *et al.*, 2023).

2. Drawbacks of the composite equivalence index

2.1 The composite equivalence index lacks discriminatory power

When introducing a new metric, researchers must offer conceptual and empirical evidence that its design and the suggested thresholds are able to differentiate between different outcomes. For example, proposing a threshold that is designed to preferentially produce one specific outcome over another does not provide any useful guidance for researchers. Unfortunately, the proposal to adopt a CEI threshold of 0.95 is not backed up by relevant conceptual arguments or formal tests that compare equal weights and differentiated weights systematically. If the same approach was to be followed, a researcher could arbitrarily declare any level of construct correlation as indicative of discriminant validity or lack thereof – a practice which has long been viewed as scientifically inappropriate. For example, methodologists have proposed formal tests (Jöreskog, 1971) or standards of comparison (Fornell and Larcker, 1981) to establish discriminant validity rather than leaving it to subjective assessments of which correlations may be deemed "conservative" or "too high" or "too low" without any further elaboration or conceptual argument (e.g. Buckley and Voorhees, 2017; Bagozzi *et al.*, 1991; McNeish and Wolf, 2023).

In this vein, McNeish's (2023) findings raise serious doubts about the appropriateness of the proposed CEI cutoff of 0.95. Using a simulation study, the author explored the psychometric properties of scores computed by equal versus differentiated weights and found that differentiated weights produced scores that had:

- higher correlations with true scores;
- higher sensitivity; and
- higher reliability than sumscores even when the two scores were correlated as highly as 0.98.

Thus, even correlations as high as 0.98 do not provide a guarantee that equal weights will perform equally well as differentiated weights.

The situation is further complicated by the fact that researchers using a weighting method like PLS-SEM can expect such high correlations. This is because the reliability and validity assessments in reflective measurement models require highly correlated indicators whose loadings meet the common 0.70 threshold, suggesting that the construct explains at least 50% of each indicator's variance (Hair *et al.*, 2022, Chap. 4; Sarstedt *et al.*, 2021). At the same time, indicator loadings should not be too high as this indicates semantic redundancy,

which might result in undesirable response patterns (Diamantopoulos *et al.*, 2012) and induce error term correlations (Drolet and Morrison, 2001). Because of these boundaries, the (correlation) weights used to compute the composite scores generally have little variation, which makes the differences between equal weights and differentiated weights less pronounced (Cohen, 1990) and will automatically translate into high CEI values. Note again that this does not imply that equal weights and differentiated weights will produce the same results (McNeish and Wolf, 2020; McNeish, 2023). The CEI proposal goes a step further to include situations where the analysis produces low values (<0.95) for individual constructs in a model. In this case, the recommendation is that researchers may use differentiated PLS-SEM weights only if they can provide theory-based expectations as to why they expect "substantial differences" in the item weights. If no such justification can be offered, researchers should revert to equal weights estimation. The problem with this recommendation is that such substantial differences can hardly be justified theoretically, as doing so would run contrary to the central tenets in the domain sampling model underlying reflective measurement, which assumes that all indicators are essentially interchangeable (Churchill, 1979). Hence, the CEI threshold of 0.95 along with the requirements to offer theory-based justification for differentiated weights for lower CEI values, is practically a one-way street to equal weights estimation.

While the above discussion relates to reflective measurement models, the use of the CEI is no less problematic in formative models. As is widely known, indicator weightings play a prominent role in formative measurement models where indicators are not necessarily highly correlated and different weights are to be expected. Yet, even in formative models, the CEI and its 0.95 threshold lack discriminatory power. Consider, for example, the extended corporate reputation model used in Hair *et al.* (2014, 2017a, 2022), which also includes four formatively measured constructs. Estimating the model using the data set of mobile network operators ($n = 344$) provided by Hair *et al.* (2022), and using the SmartPLS software (Ringle *et al.*, 2022) [3], identifies the (formatively specified) construct *Quality* as having the strongest total effect on the model's final target construct, *Customer Loyalty* [4]. The results also show that the indicators' relative contributions to forming *Quality* vary considerably (Cenfetelli and Bassellier, 2009). Specifically, whereas the indicator weights of *qual2* (0.041), *qual3* (0.106) and *qual4* (-0.005) are not statistically significant, *qual6* makes a significant ($p < 0.05$) and meaningful (0.398) relative contribution to forming *Quality*. The indicator weights of the other formatively specified constructs show similar patterns (Figure 1(a); see also Hair *et al.*, 2022). Despite these considerable differences, computing the CEI for these constructs produces values of 0.962 for *Attractiveness*, 0.960 for *Corporate Social Responsibility*, 0.983 for *Performance* and 0.972 for *Quality* [5]. If one were to consider the 0.95 CEI threshold as a meaningful standard, then all four formative constructs would be estimated using equal weights [6]. Doing so, however, would substantially diminish the theoretical and practical relevance of results in Figure 1, as we discuss below.

To illustrate this point, consider Figure 2, which shows the results of an importance performance map analysis (Ringle and Sarstedt, 2016; Slack, 1994) of the *Quality* indicators. The *x*-axis shows the indicators' standardized total effects on *Customer Loyalty* (i.e. their importance), while the *y*-axis documents the weighted and rescaled indicator averages (i.e. their performance). Using such an importance-performance map, researchers seek to identify indicators that have a particular importance for the target construct (i.e. a high total effect), but exhibit a relatively low performance (i.e. a low weighted and rescaled indicator average). However, assuming equal weights in this analysis leads to an absurdity, as these weights provide no guidance on how to prioritize their marketing activities. Aspects such as "The products/services offered by [the company] are of high quality" (*qual1*), "[The company] is
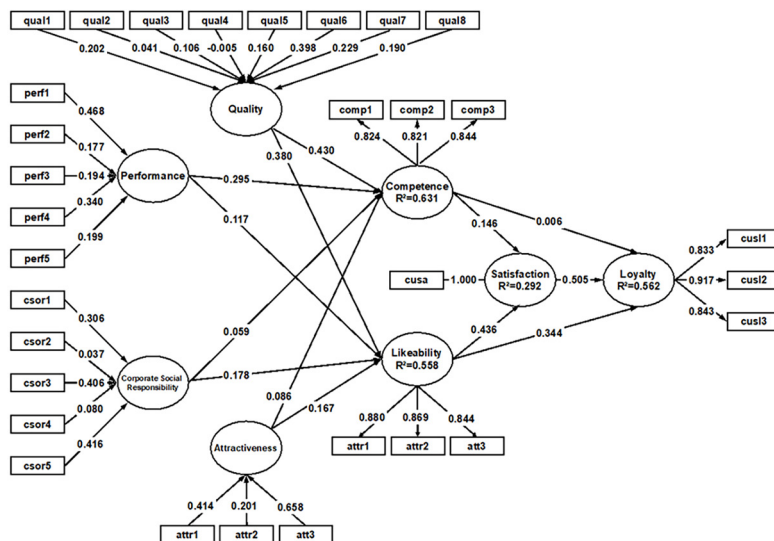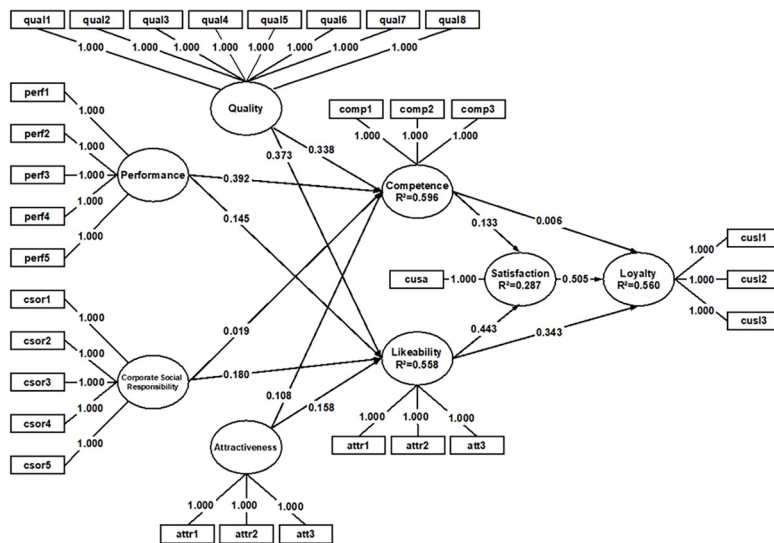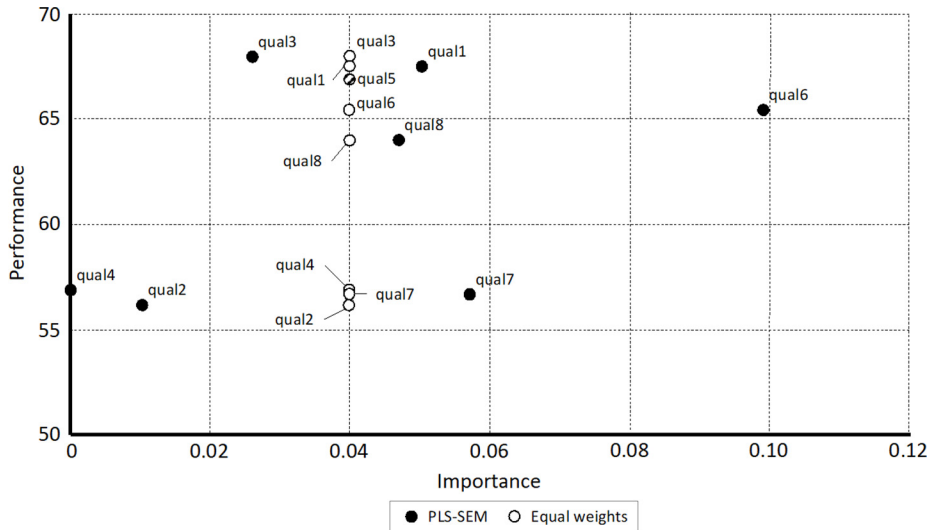
Notes: (a) PLS-SEM results; (b) equal weights results
Source: Authors' own work

Source: Authors' own work

an innovator, rather than an imitator with respect to [industry]" (*qual2*) and "[The company] is a reliable partner for customers" (*qual6*) would all be weighted equally (unfilled circles in Figure 2), even though they have vastly different practical implications for marketing communication activities. On the contrary, PLS-SEM identifies reliability aspects (*qual6*) as the strongest *Quality*-related driver of *Customer Loyalty*, while, for example, being perceived as an innovator (*qual2*) plays virtually no role (filled circles in Figure 2).

Considering the growing concerns about marketing research's relevance for business practice (Homburg *et al.*, 2015; Jaworski, 2011; Kohli and Haenlein, 2021; Kumar, 2017), discarding such additional information further removes academia from providing concrete guidance to managerial decision makers. As Jedidi *et al.* (2021, p. 22) note, "given that marketing is an applied discipline, articles published in academic journals should fulfill marketing practitioners' informational needs and be relevant to marketing practice." Offering such differentiated results is a very valuable building block in this puzzle. In other words, the knowledge of which items are strongly endorsed over others *matters* when making decisions (McNeish and Wolf, 2020).

Assuming equal indicator weights as the default not only washes out individual indicator differences and effects but also runs contrary to the principles of formative measurement model evaluation, which requires researchers to interpret each indicator's relative contribution in forming a construct (Cenfetelli and Bassellier, 2009). Using equal weights would also call into question the content validity of any index construction, as indicators that cover the entire conceptual domain *and* do not correlate too highly can hardly be assumed to have the same relative contribution in forming the construct (Diamantopoulos and Winklhofer, 2001). Nevertheless, weights can also show little variation in formative measurement models, particularly when using many indicators. The greater the number of indicators in a measurement model, the smaller the average weight (Hair *et al.*, 2022). Hence, complex measurement models will have higher CEI values by design – a characteristic that is not considered by the CEI when correlating scores computed using equal versus differentiated weights.

2.2 High composite equivalence index values can trigger substantial differences in structural model estimates
The proposed CEI guidelines imply that researchers can expect at best trivial differences in terms of structural model estimates between equal weights regression and PLS-SEM – we show that this is not the case. For example, Hair *et al.* (2017b) find that PLS-SEM achieves considerably higher levels of statistical power compared to equal weights when the effect and sample sizes are small and the indicator weights are equal in the population. Specifically, assuming an effect size of 0.15 and a sample size of $n = 100$ (and $n = 250$), the analysis produces a statistical power of 75% (and 90%) when using PLS-SEM weights, whereas equal weights estimation yielded lower statistical power of 61% (and 82%), respectively. Assuming a much simpler model, Yuan *et al.* (2020) find that PLS-SEM weights yield smaller root mean square errors in the structural model estimates than equal weights in practically all conditions, except when the exogenous construct has double the number of indicators compared to the endogenous construct.

To further illustrate the impact of assuming equal weights on structural model estimates, consider again the corporate reputation model example (Figure 1). The results indicate substantial differences in the model estimates involving the formatively measured constructs *Quality*, *Performance*, *Corporate Social Responsibility* and *Attractiveness* – despite the fact that the CEI values are higher than 0.95 in all constructs involved in the corresponding (partial) structural model regressions. For example, while the effect of *Quality* on *Competence* is 0.430 in the PLS-SEM analysis [Figure1(a)], equal weights estimation produces a considerably lower effect of 0.338 [Figure1(b)]. At the same time, equal weights estimation yields a much higher estimate for the relationship between *Performance* and *Competence* (0.392) compared to PLS-SEM (0.295). On the contrary, the path relationships involving the reflectively measured outcome constructs do not seem to be affected by the choice of scoring method in this particular case. However, as we show in the next section, equal weights estimation is still inappropriate for reflective measurement models because it conceals measurement model issues.

2.3 The composite equivalence index conceals measurement model issues
Simulation studies comparing the relative performance of PLS-SEM and equal weights generally assume the measurement models are reliable and valid. However, in empirical applications of the methods, this is rarely the case as indicators often do not meet minimum quality standards in terms of loadings in reflective measurement models or weights in formative measurement models (Sarstedt *et al.*, 2021). More specifically, equal weights can easily conceal potential reliability and validity problems. To demonstrate these problems, we draw on the European customer satisfaction index (ECSI) model and the data set used by Tenenhaus *et al.* (2005). Figure 3(a), shows the model results using PLS-SEM-based weights while Figure 3(b) shows the results of the equal weights estimation.

The analysis produces CEI values of 1 for all constructs except for *Loyalty*, which has a CEI value of 0.93 and is slightly below the proposed CEI threshold of 0.95. Since *Loyalty* is specified reflectively, researchers would generally not expect substantial differences in the indicators' correlation weights (i.e. loadings). In this case too, the logic behind the proposed CEI would suggest applying equal weights for the *Loyalty* construct. But arbitrarily imposing equal weights risks retaining measurement model indicator estimates that do not meet established minimum quality standards.
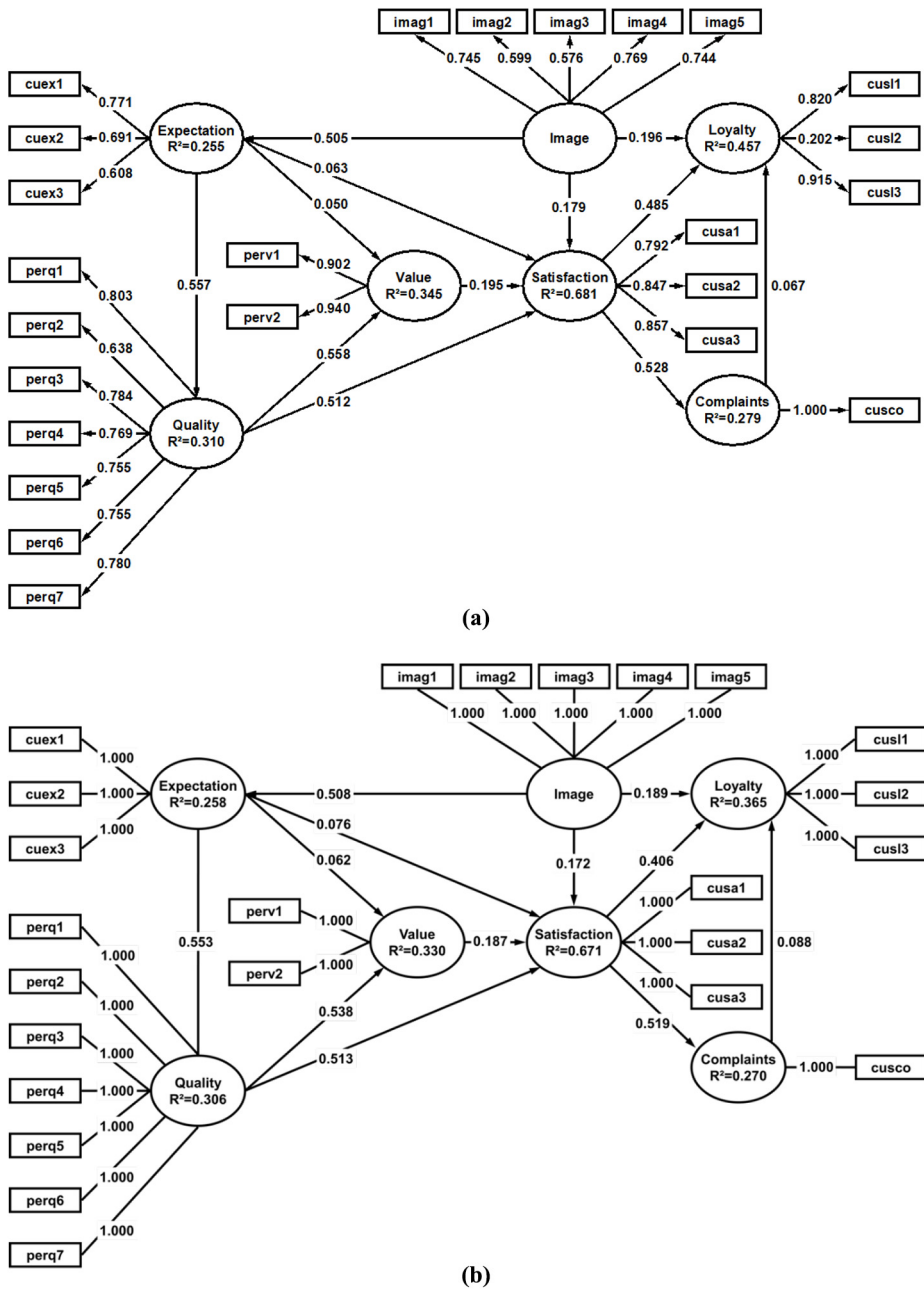
Figure 3.
ECSI model results

Notes: (a) PLS-SEM results; (b) equal weights results
Source: Authors' own work

Specifically, in this situation *Loyalty* has an unreliable indicator *cusl2*, as indicated by the low loading of 0.203. If we take a closer look at the indicator on the basis of this result, it quickly becomes clear that the ambiguity in the survey question (respondents were asked to indicate the percentage price difference at which they would switch to a competitor) is the reason for the unreliability of this indicator (see Table 1 in Tenenhaus *et al.*, 2005). A researcher relying on equal weights would not be able to identify this problem because of the same weight assigned to all three indicators, including the unreliable indicator, thereby increasing the degree of random error in the *Loyalty* construct. This increase in random error results in a considerably lower R^2 value in the equal weights estimation ($R^2 = 0.365$) compared to the PLS-SEM results ($R^2 = 0.457$) – see Figure 3. Similarly, the random error deflates path coefficient estimates of relationships pointing at the *Loyalty* construct; for example, the relationship from *Satisfaction* to *Loyalty* drops from 0.485 to 0.406.

To further illustrate the effectiveness of PLS-SEM weights, consider an alternative model in which the unreliable indicator *cusl2* is excluded from the model. Estimating this reduced model [Figure 4(a)] shows that the PLS-SEM results are almost identical to those produced by the PLS-SEM estimation of the model *with* the unreliable indicator [Figure 3(a)]. In contrast, the same analysis using equal weights estimation produces obviously different results. For example, assuming equal weights for the model with the unreliable indicator produces an estimate of 0.406 for the relationship between *Satisfaction* and *Loyalty* [Figure 4 (b)], whereas the same effect is estimated at 0.464 in the model without the unreliable indicator [Figure 4(b)]. Thus, relying on equal weights not only causes researchers to lose the ability to identify unreliable indicators but their results also become more likely to be negatively affected by the accidental inclusion of such indicators.

To summarize, relying on the CEI can cause researchers to easily overlook the problems caused by the unreliable indicators – such as *cusl2* in the ECSI example above. The corresponding results not only bias the structural model estimates but also can easily cause Type II errors (i.e. false negatives), triggered by the deflation of path coefficient estimates. In the Appendix, we report two additional examples that illustrate the problems arising from CEI's concealing of measurement model issues.

Furthermore, assessing the measurement model quality based on the PLS-SEM results but then using sumscores for the final structural model estimation would detach the validity assessment from the scores, which is in sharp contrast to common psychometric standards (McNeish, 2023). For example, Borsboom *et al.* (2003, pp. 206–207) noted that "the assumption that it was this model, and not some other model, that generated the data must precede the estimation process. In other words, if one considers the weighted sumscore as an estimate of the position of a given subject on a latent variable, one does so under the model specified." Or as McNeish (2023, p. 4271) noted, "if sum scores are used to draw inferences, then the validity assessment should be based on a model that is consistent with sum scoring rather than a separate factor model" (i.e. a model that assumes differentiated weights). Thus, the application of the CEI could easily produce misleading conclusions both at the measurement model and structural model levels.

2.4 The use of equal weights and composite equivalence index leads to inferior out-of-sample prediction accuracy

While several studies have compared equal weights and PLS-SEM estimation in terms of bias (e.g. Hair *et al.*, 2017b; Yuan *et al.*, 2020), only Becker *et al.* (2013) compared the two methods in terms of predictive power. Using a simple model with two exogenous and one
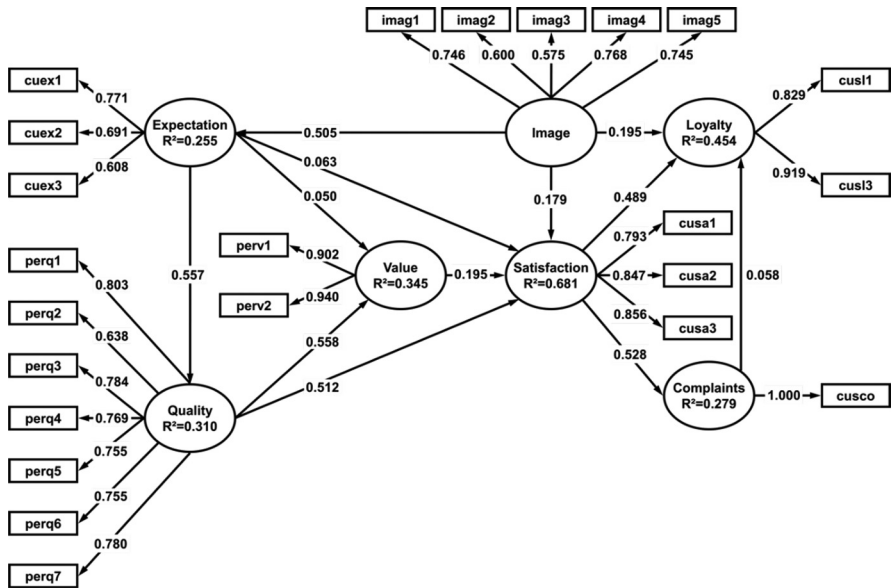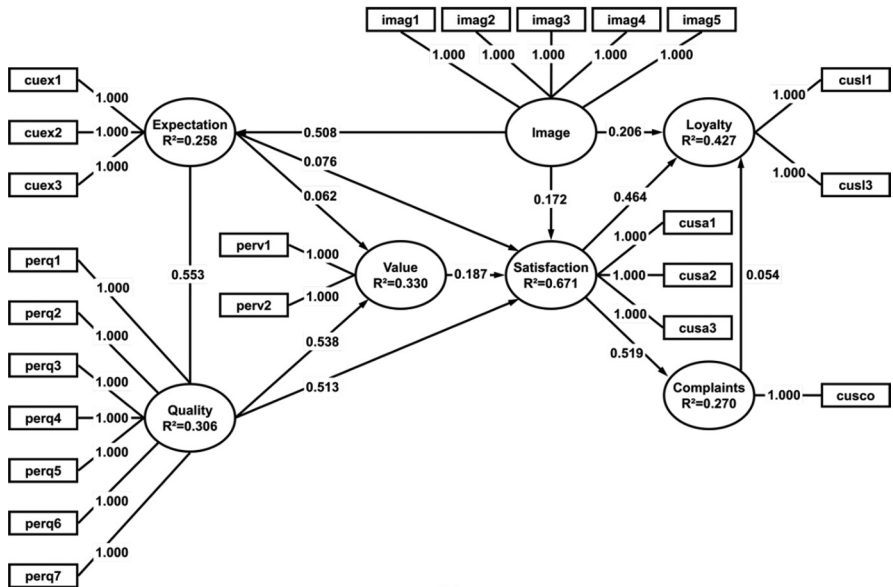
Figure 4.
ECSI model results
without *cusl2*

Notes: (a) PLS-SEM results; (b) equal weights results
Source: Authors' own work

endogenous construct, they show that relying on PLS-SEM weights achieves higher levels of predictive power than on equal weights, except in situations with small sample sizes. Extending Becker *et al.* (2013), we conduct a more comprehensive simulation study to compare PLS-SEM-based differentiated weight's predictive capabilities compared with those of equal weights – especially in situations where CEI recommends the use of equal weights.

Our simulation study relies on the population model (Figure 5) used by Hair *et al.* (2017b), which closely resembles model set-ups routinely used in empirical research. Unlike Hair *et al.* (2017b), however, we do not include misspecified paths in our study as our objective is not to assess Type I error rates (i.e. false positives). When generating the data, we assume path coefficients of 0.4 for all structural model relationships. For the measurement model, we keep the value of 0.8 fixed for the correlation weights related to the white indicators in Figure 5 (e.g. $x_{1,1}$ and $x_{2,1}$), but we systematically decrease those for the grey-shaded ones (e.g. $x_{3,1}$ and $x_{4,1}$) from 0.8 by steps of 0.1 until 0 has been reached. That is, we increase the difference in correlation weights between the grey-shaded indicators and the white indicators in steps of 0.1 from 0 to 0.8. We then generate data for all nine factor levels using a sample size of 500, which provides reasonably stable results (e.g. Reinartz *et al.*, 2009) and is sufficiently large for a model with the complexity shown in Figure 5 (Hair *et al.*, 2022, Chap. 1). For each factor level we next run 5,000 replications.

We use PLS-SEM-based differentiated weights and equal weights methods to estimate the model for all simulation runs and compute the CEI values. Our simulation study uses operative prediction to calculate the out-of-sample prediction value (i.e. predictions from exogenous indicators to endogenous indicators; Shmueli *et al.*, 2016; Danks, 2021). To compute the out-of-sample prediction errors for each estimation method, we use a large validation data set with 10,000 observations (see also van Smeden *et al.*, 2019, who use a
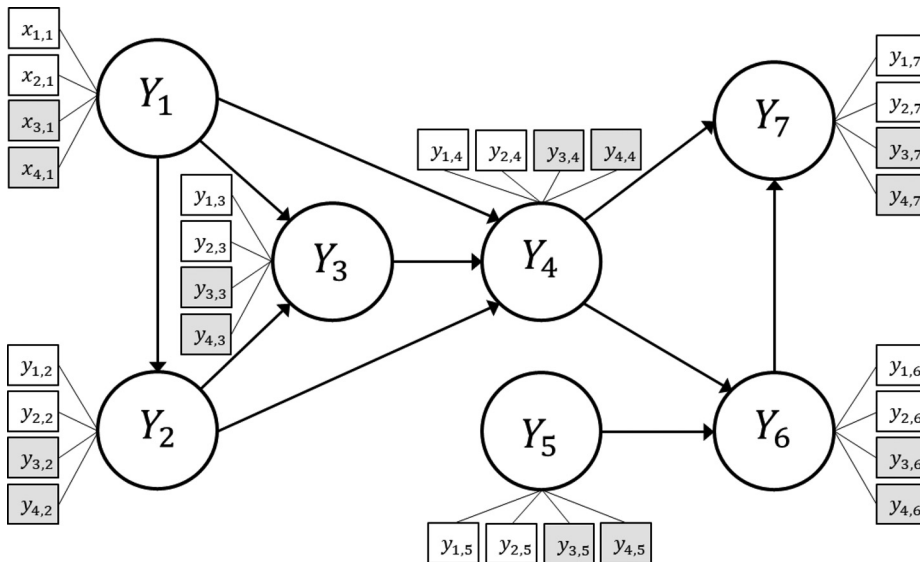

Figure 5.
Population model

Source: Authors' own work

validation data set with 5,000 observations). Finally, we apply the squared prediction error metric to evaluate the out-of-sample prediction performance of PLS-SEM and equal weights (Hastie *et al.*, 2009).

The simulation results show that for both the PLS-SEM weights and equal weights methods, the lowest squared prediction error (i.e. best predictive accuracy) occurs for Y_4, while the highest squared prediction error (worst predictive accuracy) occurs for Y_2. Therefore, we focus on these two constructs in the presentation of results in Table 1 (note that the results of all other constructs lie in between those of Y_2 and Y_4). As can be seen, equal weights estimation produces better out-of-sample prediction accuracy (i.e. lower squared prediction errors in the majority of simulation runs) than PLS-SEM weights in exactly *one* specific – and unrealistic – condition: when the population model consists of equal weights *for all indicators in all of the constructs* (Table 1; see the first

| Construct	Indicators[a]	Correlation weights	Difference in indicators' correlations weights[b]	Percentage of CEI values above the 0.95 threshold (%)	Percentage of simulation runs where PLS-SEM weights offer superior predictions compared to equal weights (%)
Y_2	$y_{1,2}$	0.8	0	100	45.7
	$y_{4,2}$	0.8			48.0
	$y_{1,2}$	0.8	0.1	100	82.8
	$y_{4,2}$	0.7			84.4
	$y_{1,2}$	0.8	0.2	100	96.1
	$y_{4,2}$	0.6			96.3
	$y_{1,2}$	0.8	0.3	100	99.2
	$y_{4,2}$	0.5			98.5
	$y_{1,2}$	0.8	0.4	100	99.8
	$y_{4,2}$	0.4			98.6
	$y_{1,2}$	0.8	0.5	79.2	99.8
	$y_{4,2}$	0.3			97.8
	$y_{1,2}$	0.8	0.6	5.3	99.9
	$y_{4,2}$	0.2			97.3
Y_4	$y_{1,4}$	0.8	0	100	34.6
	$y_{4,4}$	0.8			34.0
	$y_{1,4}$	0.8	0.1	100	100
	$y_{4,4}$	0.7			99.4
	$y_{1,4}$	0.8	0.2	100	100
	$y_{4,4}$	0.6			100
	$y_{1,4}$	0.8	0.3	100	100
	$y_{4,4}$	0.5			100
	$y_{1,4}$	0.8	0.4	100	100
	$y_{4,4}$	0.4			100
	$y_{1,4}$	0.8	0.5	85.2	100
	$y_{4,4}$	0.3			100
	$y_{1,4}$	0.8	0.6	2.6	100
	$y_{4,4}$	0.2			99.9

Notes: [a]We report the results of indicators 1 and 4 in this table; the results of indicator 2 are similar to those of indicator 1 and the results of indicator 3 are similar to those of indicator 4. [b]The correlation weights of indicators 1 and 2 are fixed at 0.8. The correlation weights of indicators 3 and 4 systematically obtain lower values so that the difference of these indicators increases. The column shows the difference between the correlation weights of indicators 1 and 2 and indicators 3 and 4 per composite
Source: Authors' own work

Table 1.
Simulation results

two rows of Y_2 and Y_4 for indicators 1 and 4 with a correlation weight of 0.8). This is not surprising because equal weights estimation exactly matches the data generation model for this specific condition. However, in line with McNeish and Wolf (2020), we note that such a condition (i.e. equal weights for all indicators) is very unlikely to exist in real conditions and practical data sets. In contrast, under the realistic conditions where differences in the indicator weights are expected to exist, the picture clearly changes in that PLS-SEM weights almost always outperform equal weights in predictive capabilities – even when the differences in the indicator weights are marginal. For example, when the difference in indicator weights is merely 0.1, PLS-SEM weights show better prediction than equal weights in 82.8% (Y_2) to 100% (Y_4) of all simulations. As Table 1 shows, higher indicator weights differences lead to more pronounced out-of-sample prediction advantages for PLS-SEM weights and clearly reveals the inferiority of equal weights.

To analyze the behavior in more detail, we contrast the out-of-sample mean squared error for PLS-SEM weights and equal weights estimations for various levels of correlation weight differences. Figure 6 shows these results for the indicators $y_{1,2}$ and $y_{1,4}$ (i.e. the indicators with fixed weights) and indicators $y_{4,2}$ and $y_{4,4}$ (i.e. the indicators with varying weights) of Y_2 and Y_4. The results show that for increasing differences of the indicators' correlation weights, the out-of-sample predictive power of $y_{4,2}$ and $y_{4,4}$ (i.e. the indicators whose correlation weights are being lowered) decreases [see the upper lines in Figure 6(a) and (b)]. For the indicators with increasingly smaller correlation weights, we find a small advantage for PLS-SEM weights (i.e. a lower sum of squared prediction errors compared with equal weights). More interestingly, for the indicators with a fixed correlation weight ($y_{1,2}$ and $y_{1,4}$), the increase of the sum of squared prediction error is less pronounced [see the lower lines in Figure 6(a) and (b)].

(a)   **(b)**
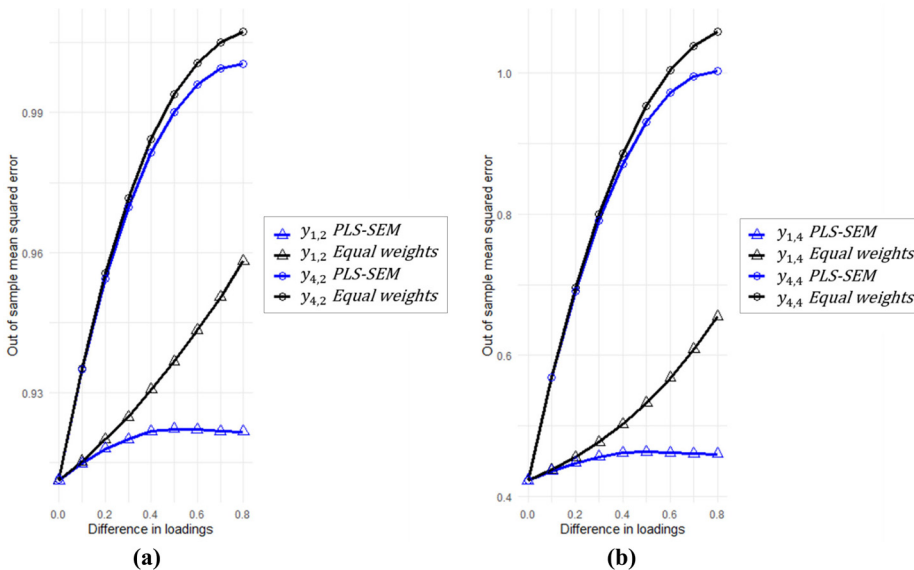
Notes: (a) *Y2*; (b) *Y4*
Source: Authors' own work

Figure 6.
Squared prediction
errors

However, while the PLS-SEM weights' prediction quality seems not to be affected by the varying the weight differences, the prediction error of equal weights substantially increases for these indicators (i.e. increasingly worse prediction accuracy).

These results clearly show that PLS-SEM weights have higher predictive power in the vast majority of cases – yet, the CEI still recommends that researchers should use equal weights for their model estimation. Thus, relying on the CEI can misguide researchers into choosing the equal weights method that offers inferior predictive accuracy. Specifically, only when the differences in the indicator weights reach 0.6 does the CEI fall below the threshold of 0.95 in most of the simulation runs. In all other situations, the CEI is above 0.95 in 79.2%–100% of simulation runs (Table 1). These results raise severe concerns about the CEI-based recommendation that researchers should use equal weights for the model estimation even when there are clear sizable differences in indicator weights in the population model. Thus, reliance on CEI is likely to take the researcher away from the population model and lead to inferior predictive accuracy – an important goal in PLS-SEM analyses.

To further illustrate the implications on out-of-sample predictive performance when using CEI, we revisit the corporate reputation example. To broaden the analysis, we include an additional data set from Sarstedt *et al.*'s (2023) conceptual replication of the extended corporate reputation model (Figure 1) [7]. In the following, we refer to the data set used in Hair *et al.* (2022) ($n = 344$) as *Corprep22*, whereas the new data set ($n = 308$) will be referred to as *Corprep23*. To compare the out-of-sample predictive performance of equal weights and PLS-SEM weights, we use the cross-validated predictive ability test (CVPAT; Liengaard *et al.*, 2021; Sharma *et al.*, 2023) with the statistical software R (R Core Team, 2022). The CVPAT evaluates if the loss (i.e. the average squared out-of-sample prediction error) differs between two models – in our case models estimated with PLS-SEM-based and equal weights. The model with the lowest loss is considered to be the best in predicting new observations. As in Sharma *et al.* (2023), we focus on the combined predictive ability of *CUSA* and *CUSL* as these constructs are particularly relevant for drawing managerial implications. We note that because *CUSA* is a single-item construct, PLS-SEM and the equal weights approach will have the same measurement model for this construct. Furthermore, the indicators of *CUSL* have similar loadings in both data sets, which, in turn, will make the weights close to the equal weights scenario. In both the *Corprep22* and *Corprep23* data sets, the CEI for *CUSL* is 0.999, approaching the theoretical maximum value of 1 as closely as can be reasonably expected. Despite these high CEI values, we find that PLS-SEM has a better out-of-sample prediction accuracy than equal weights in both data sets as evidenced in lower loss values (see Table 2). While the difference in loss values is not significant for the *Corprep22* data set, for the *Corprep23* data set, PLS-SEM's predictive accuracy is significantly better ($p < 0.01$) than that of equal weights.

Data set	PLS-SEM loss	Equal weight loss	Difference	*p*-value
Corprep22	0.629	0.630	−0.001	0.493
Corprep23	0.555	0.559	−0.004	0.003

Notes: Loss = average loss. The calculation follows the $CVPAT^{compare}_{construct}$ measure from Sharma *et al.* (2023), but instead of testing the out-of-sample predictive ability between two PLS-SEM, the out-of-sample predictive ability is tested between a PLS-SEM and an equal weight model. *p*-values are based on 5,000 bootstrap samples
Source: Authors' own work

Table 2.
CVPAT results for *CUSA* and *CUSL*

3. Discussion

Following on the recent call for investigation (Voss, 2023), the purpose of our paper was to shed light on the proposal to adopt equal weights (sumscores) as the default scoring method and the corresponding use of CEI in the PLS-SEM framework–and why we believe it is incorrect. While we assessed equal weights estimation and the CEI in the specific context of PLS-SEM, many of the problems associated with sumscores that are highlighted in this article are relevant to other composite-based methods as well (e.g. GSCA). There seem to be two factors behind the justification to use sumscores:

(1) they are easy to compute; and
(2) they are generally highly correlated with weighted scores, thereby potentially providing similar results.

Both arguments stand on quicksand. First, while the ease of calculating sumscores had a useful role to play in the predigital era when computing resources were at the premium, this is no longer the case. Researchers now have access to advanced software that produce differentiated scores by default, and with ease. In this light, creating sumscores and then using them as inputs to run partial regression seems like jumping through "hoops to justify broadly recommending classic but beatable methods that were developed in a vastly different computational landscape" (McNeish, 2023, p. 4280). Second, while a correlation of 0.95 or 0.98 may seem extremely high to empirical researchers, "correlations between estimands in statistical research are routinely very high because different methods are repackaging identical information" (McNeish, 2023, p. 4280). However, a high correlation (even as high as 0.98) between sumscores and weighted scores does not mean that there will be nontrivial differences in the results produced by the two scoring methods (McNeish, 2023). In particular, prior research and our simulation results show that the proposed CEI cutoff value (0.95) is not only inappropriate but also that *any* cutoff value that results in the use of equal weights estimation over differentiated weights will lead to conceptual and empirical issues (Andersson and Yang-Wallentin, 2020; Lastovicka and Thamodaran, 1991; Estabrook and Neale, 2013; Hair *et al.*, 2017b; Murray *et al.*, 2016; Rigdon *et al.*, 2019; McNeish and Wolf, 2020; McNeish, 2022, 2023). The totality of this evidence shows that Voss' (2023) call for "adequate cutoffs for agreement" is, unfortunately, infeasible. If equal weights can only perform empirically as well as differentiated weights in very specific conditions, and no better in most practical conditions, then there are solid reasons to be doubtful of their application (McNeish, 2023).

In addition to previous accounts of the relative disadvantages of equal weights as compared to differentiated weights (e.g. Murray *et al.*, 2016; Estabrook and Neale, 2013; McNeish and Wolf, 2020; McNeish, 2023; Hair *et al.*, 2017b), our assessment of CEI shows that it suffers from major weaknesses. We find that it lacks discriminatory power, conceals reliability concerns in reflective measurement models, overlooks the likelihood of differences in relative indicator contributions in formative measurement models and violates the principles of index construction procedures. Moreover, our simulation results show that equal weights exhibit inferior out-of-sample predictive power compared to PLS-SEM weights. Yet, the CEI suggested using equal weights in practically all instances in our simulation, thereby recommending a demonstratively inferior scoring method whose disadvantages have been broadly discussed in prior literature (e.g. McNeish and Wolf, 2020; Lastovicka and Thamodaran, 1991; Andersson and Yang-Wallentin, 2020; McNeish, 2023). In sum, adoption of the CEI and equal weights would decrease the methodological rigor of research.

Furthermore, adopting CEI and equal weights is inconsistent with the focus of PLS-SEM's causal-predictive nature. For example, Hair *et al.* (2019, p. 3) note that PLS-SEM "emphasizes prediction in estimating statistical models, whose structures are designed to provide causal explanations [and] thereby overcomes the apparent dichotomy between explanation—as typically emphasized in academic research—and prediction, which is the basis for developing managerial implications." Sarstedt and Danks (2022) recently stressed the need to emphasize prediction in the estimation of models that have been derived from theory and logic. A well-fitting model in an exploratory power sense does not necessarily perform well in terms of prediction, yet its predictive accuracy determines its practical relevance to a large extent. In light of these developments, adopting a metric like the CEI that univocally favors an estimator that lags behind in terms of predictive power is highly problematic.

Several other arguments clearly speak against the adoption of equal indicator weights even when the correlation between scores from equal and differentiated indicator weights is close to unity – as convincingly laid out by McNeish and Wolf (2020) and McNeish (2023). For example, assuming differentiated weights facilitates the coherent application of reliability and validity metrics to safeguard the quality of results and allows for more differentiated conclusions to be drawn. Adopting differentiated weights also avoids enforcing unnatural constraints on the empirical model, which have not been imposed in prior applications of established scales that researchers adopt in their follow-up studies. Because sumscores assume that all items provide the same amount of information across all respondents and populations, they assume strict measurement invariance across samples, and preclude the possibility that items may work differently in different populations (McNeish, 2023). For example, Schlägel and Sarstedt (2016) found substantial differences in indicators measuring the various dimensions of cultural intelligence across respondents from Germany, Turkey and the USA, showing that measurement invariance should not be assumed. The same analysis using sumscores, however, would have established measurement invariance by design, yielding potentially questionable results in follow-up multigroup comparisons. In this vein, McNeish (2022, p. 200) notes that the "challenge of working in behavioral research is that many variables are unobservable constructs. Any set of item responses can be added together to form a score, but such scores are not necessarily meaningful. Measurement models and psychometrics provide necessary evidence that the items measure a single construct and that the item responses can be interpreted as manifestations of the unobservable construct. In absence of this evidence, the connection between item responses and the intended construct is unclear and results from studies lacking validity information are ambiguous at best. . . ."

To summarize, the use of CEI and equal weights would adversely affect the validity of measurement and structural model results and yield models with inferior out-of-sample predictive power. Nevertheless, any effort to develop measures that identify situations that call for preferring equal weights over differentiated indicator weights as produced by PLS-SEM is highly welcome. Given the various lenses through which the equal weights and differentiated weights can be compared – bias and predictive power being only two of them – this is certainly a challenging undertaking that goes well-beyond contrasting composite scores. For example, Wilks (1938) showed that the performances of unweighted and weighted scores converge as the number of items increases unboundedly. However, McNeish (2023) found that this aspect does not arise even when the number of items reaches 15 for reflective models. The lower bound of where this convergence emerges in practical conditions is not yet known – and McNeish's (2023) results suggest that this bound may be greater than 15 items per construct, a condition hardly encountered in PLS-SEM research

(e.g. because of violations of unidimensionality). More research is needed to ascertain this lower bound, especially in composite models.

Finally, our empirical contrasting of the predictive accuracy of differentiated and equal weights has shown that PLS-SEM-based weights are much better suited for one of the data sets, but not the other. In light of these results, a potential avenue for future research could involve investigating whether the comparative predictive performance of PLS-SEM versus equal weights can serve as a metric to assess the compatibility of the data with a specific model.

4. Methodological implications

Our results show that the CEI lacks discriminatory power and its use produces distorted results:

- with major differences in structural model estimates;
- that conceal measurement model issues; and
- almost always result in inferior out-of-sample predictive accuracy, which is of central concern in PLS-SEM analyses (Danks *et al.*, 2023).

Relatedly, given the ample evidence in favor of differentiated indicator weights over equal weights on both conceptual (e.g. McNeish and Wolf, 2020) and empirical grounds (e.g. Yuan *et al.*, 2020), and the results discussed in this paper, we recommend against the use equal weights as the preferred methods of choice under most conditions. Instead, researchers should rely on differentiated indicator weights such as those provided by PLS-SEM or other composite-based SEM methods. If researchers nevertheless decide to use equal weights estimation, we recommend that they provide strong ex ante theoretical justification for their assumption why each indicator or item contributes an equal amount of information to the composite construct in the population.

Notes

1. In the following we will use the terms "equal weights," "unit weights," and "sumscores" interchangeably to refer to the method where all measurement model indicators are constrained to have same weights and summed to produce composite scores.

2. Note that the CEI has been discussed in the context of PLS-SEM, but is applicable to any other composite-based SEM method without limitations.

3. The data set and SmartPLS project files can be downloaded from www.pls-sem.net/downloads/3rd-edition-a-primer-on-pls-sem-1/. We estimated the model using the standard PLS-SEM algorithm, assuming unit weights in the initialization and using the path weighting scheme for model estimation. We applied mean replacement for treating missing values. See Hair *et al.* (2021, 2022), and Sarstedt *et al.* (2021) for details regarding the model estimation.

4. Specifically, the total effects of the four antecedent constructs are 0.248 (*Quality*), 0.105 (*Corporate Social Responsibility*), 0.101 (*Attractiveness*) and 0.089 (*Performance*).

5. Computing the CEI values requires extracting the final sets of constructs scores from the model estimation with equal weights and differentiated weights (e.g. obtained through PLS-SEM). For example, in the SmartPLS 4 output, the construct scores can be accessed in the results report under *Final results* ➔ *Latent variables* ➔ *Scores*. Researchers then need to correlate the vectors of sumscores and weighted scores for each construct to obtain the construct-specific CEI.

6. Note that this result is not grounded in potentially misspecified measurement models as the findings from a confirmatory tetrad analysis (Gudergan *et al.*, 2008) and measurement-theoretic

considerations clearly speak in favor of the formative specification – see Hair *et al.* (2024, Chap. 3) for details.

7. The data set and SmartPLS project files can be downloaded from https://osf.io/6hdxn/

References

Andersson, G. and Yang-Wallentin, F. (2020), "Generalized linear factor score regression: a comparison of four methods", *Educational and Psychological Measurement*, Vol. 81 No. 4, pp. 617-643.

Bagozzi, R.P., Yi, Y. and Philipps, L.W. (1991), "Assessing construct validity in organizational research", *Administrative Science Quarterly*, Vol. 36 No. 3, pp. 421-458.

Becker, J.-M., Rai, A. and Rigdon, E.E. (2013), "Predictive validity and formative measurement in structural equation modeling: embracing practical relevance", *Proceedings of the International Conference on Information Systems*, Milan.

Bobko, P., Roth, P.L. and Buster, M.A. (2007), "The usefulness of unit weights in creating composite scores: a literature review, application to content validity, and meta-analysis", *Organizational Research Methods*, Vol. 10 No. 4, pp. 689-709.

Borsboom, D., Mellenbergh, G.J. and van Heerden, J. (2003), "The theoretical status of latent variables", *Psychological Review*, Vol. 110 No. 2, pp. 203-219.

Buckley, C. and Voorhees, E.M. (2017), "Evaluating evaluation measure stability", *ACM SIGIR Forum*, Vol. 51 No. 2, pp. 235-242.

Cenfetelli, R.T. and Bassellier, G. (2009), "Interpretation of formative measurement in information systems research", *MIS Quarterly*, Vol. 33 No. 4, pp. 689-708.

Churchill, G.A. (1979), "A paradigm for developing better measures of marketing constructs", *Journal of Marketing Research*, Vol. 16 No. 1, pp. 64-73.

Cohen, J. (1990), "Things I have learned (so far)", *American Psychologist*, Vol. 45 No. 12, pp. 1304-1312.

Danks, N. (2021), "The piggy in the middle: the role of mediators in PLS-SEM-based prediction", *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, Vol. 52 No. SI, pp. 24-42.

Danks, N.P., Ray, S. and Shmueli, G. (2023), "The composite overfit analysis framework: assessing the out-of-sample generalizability of construct-based models using predictive deviance, deviance trees, and unstable paths", *Management Science*, Vol. 70 No. 1, pp. 647-669.

Diamantopoulos, A. and Winklhofer, H.M. (2001), "Index construction with formative indicators: an alternative to scale development", *Journal of Marketing Research*, Vol. 38 No. 2, pp. 269-277.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P. and Kaiser, S. (2012), "Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective", *Journal of the Academy of Marketing Science*, Vol. 40 No. 3, pp. 434-449.

Drolet, A.L. and Morrison, D.G. (2001), "Do we really need multiple-item measures in service research?", *Journal of Service Research*, Vol. 3 No. 3, pp. 196-204.

Estabrook, R. and Neale, M. (2013), "A comparison of factor score estimation methods in the presence of missing data: reliability and an application to nicotine dependence", *Multivariate Behavioral Research*, Vol. 48 No. 1, pp. 1-27.

Fornell, C.G. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Gudergan, S.P., Ringle, C.M., Wende, S. and Will, A. (2008), "Confirmatory tetrad analysis in PLS path modeling", *Journal of Business Research*, Vol. 61 No. 12, pp. 1238-1249.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2014), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, Thousand Oaks, CA.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2017a), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd ed., Sage, Thousand Oaks, CA.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2022), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3 ed., Sage, Thousand Oaks, CA.

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), "When to use and how to report the results of PLS-SEM", *European Business Review*, Vol. 31 No. 1, pp. 2-24.

Hair, J.F., Sarstedt, M., Ringle, C.M. and Gudergan, S.P. (2024), *Advanced Issues in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd ed., Sage, Thousand Oaks, CA.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M. and Thiele, K.O. (2017b), "Mirror, mirror on the wall: a comparative evaluation of composite-based structural equation modeling methods", *Journal of the Academy of Marketing Science*, Vol. 45 No. 5, pp. 616-632.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P. and Ray, S. (2021), *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*, Springer, Cham.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York, NY.

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. and Calantone, R.J. (2014), "Common beliefs and reality about partial least squares: comments on Rönkkö & Evermann (2013)", *Organizational Research Methods*, Vol. 17 No. 2, pp. 182-209.

Homburg, C., Vomberg, A., Enke, M. and Grimm, P.H. (2015), "The loss of the marketing department's influence: is it happening? And why worry?", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 1-13.

Jaworski, B.J. (2011), "On managerial relevance", *Journal of Marketing*, Vol. 75 No. 4, pp. 211-224.

Jedidi, K., Schmitt, B.S., Sliman, M.B. and Li, Y. (2021), "R2M index 1.0: assessing the practical relevance of academic marketing articles", *Journal of Marketing*, Vol. 85 No. 5, pp. 22-41.

Jöreskog, K.G. (1971), "Simultaneous factor analysis in several populations", *Psychometrika*, Vol. 36 No. 4, pp. 409-426.

Kline, R.B. (2005), *Principles and Practice of Structural Equation Modelling*, Guilford Press, New York, NY.

Kohli, A.K. and Haenlein, M. (2021), "Factors affecting the study of important marketing issues: Implications and recommendations", *International Journal of Research in Marketing*, Vol. 38 No. 1, pp. 1-11.

Kumar, V. (2017), "Integrating theory and practice in marketing", *Journal of Marketing*, Vol. 81 No. 2, pp. 1-7.

Lastovicka, J.L. and Thamodaran, K. (1991), "Common factor score estimates in multiple regression problems", *Journal of Marketing Research*, Vol. 28 No. 1, pp. 105-112.

Liengaard, B.D., Sharma, P.N., Hult, G.T.M., Jensen, M.B., Sarstedt, M., Hair, J.F. and Ringle, C.M. (2021), "Prediction: coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling", *Decision Sciences*, Vol. 52 No. 2, pp. 362-392.

Lohmöller, J.-B. (1989), *Latent Variable Path Modeling with Partial Least Squares*, Physica, Heidelberg.

McNeish, D. (2022), "Limitations of the sum-and-alpha approach to measurement in behavioral research", *Policy Insights from the Behavioral and Brain Sciences*, Vol. 9 No. 2, pp. 196-203.

McNeish, D. (2023), "Psychometric properties of sum scores and factor scores differ even when their correlation is 0.98: a response to Widaman and Revelle", *Behavior Research Methods*, Vol. 55 No. 8, pp. 4269-4290.

McNeish, D. and Wolf, M.G. (2020), "Thinking twice about sum scores", *Behavior Research Methods*, Vol. 52 No. 6, pp. 2287-2305.

McNeish, D. and Wolf, M.G. (2023), "Dynamic fit index cutoffs for confirmatory factor analysis models", *Psychological Methods*, Vol. 28 No. 1, pp. 61-88.

Murray, A.L., Molenaar, D., Johnson, W. and Krueger, R.F. (2016), "Dependence of gene-by-environment interactions (GxE) on scaling: comparing the use of sum scores, transformed sum scores and IRT scores for the phenotype in tests of GxE", *Behavior Genetics*, Vol. 46 No. 4, pp. 552-572.

R Core Team (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.

Ree, M.J., Carretta, T.R. and Earles, J.A. (1998), "In top-down decisions, weighting variables does not matter: a consequence of Wilks' theorem", *Organizational Research Methods*, Vol. 1 No. 4, pp. 407-420.

Reinartz, W.J., Haenlein, M. and Henseler, J. (2009), "An empirical comparison of the efficacy of covariance-based and variance-based SEM", *International Journal of Research in Marketing*, Vol. 26 No. 4, pp. 332-344.

Rigdon, E.E. (2012), "Rethinking partial least squares path modeling: in praise of simple methods", *Long Range Planning*, Vol. 45 Nos 5/6, pp. 341-358.

Rigdon, E.E., Becker, J.-M. and Sarstedt, M. (2019), "Parceling cannot reduce factor indeterminacy in factor analysis: a research note", *Psychometrika*, Vol. 84 No. 3, pp. 772-780.

Ringle, C.M. and Sarstedt, M. (2016), "Gain more insight from your PLS-SEM results: the importance-performance map analysis", *Industrial Management and Data Systems*, Vol. 116 No. 9, pp. 1865-1886.

Ringle, C.M., Wende, S. and Becker, J.-M. (2022), *SmartPLS 4*, SmartPLS, Oststeinbek.

Rönkkö, M., Lee, N., Evermann, J., McIntosh, C.N. and Antonakis, J. (2023), "Marketing or methodology? Exposing fallacies of PLS with simple demonstrations", *European Journal of Marketing*, Vol. 57 No. 6, pp. 1597-1617.

Sarstedt, M. and Danks, N.P. (2022), "Prediction in HRM research: a gap between rhetoric and reality", *Human Resource Management Journal*, Vol. 32 No. 2, pp. 485-513.

Sarstedt, M., Ringle, C.M. and Hair, J.F. (2021), "Partial least squares structural equation modeling", in Homburg, C., Klarmann, M. and Vomberg, A.E. (Eds), *Handbook of Market Research*, Springer, Cham, pp. 1-47.

Sarstedt, M., Ringle, C.M. and Iuklanov, D. (2023), "Antecedents and consequences of corporate reputation: a dataset", *Data in Brief*, Vol. 48, p. 109079.

Schlägel, C. and Sarstedt, M. (2016), "Assessing the measurement invariance of the four-dimensional cultural intelligence scale across countries: a composite model approach", *European Management Journal*, Vol. 34 No. 6, pp. 633-649.

Sharma, P.N., Liengaard, B.D., Hair, J.F., Sarstedt, M. and Ringle, C.M. (2023), "Predictive model assessment and selection in composite-based modeling using PLS-SEM: extensions and guidelines for using CVPAT", *European Journal of Marketing*, Vol. 57 No. 6, pp. 1662-1677.

Shmueli, G., Ray, S., Velasquez Estrada, J.M. and Chatla, S.B. (2016), "The elephant in the room: evaluating the predictive performance of PLS models", *Journal of Business Research*, Vol. 69 No. 10, pp. 4552-4564.

Slack, N. (1994), "The importance-performance matrix as a determinant of improvement priority", *International Journal of Operations and Production Management*, Vol. 14 No. 5, pp. 59-75.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M. and Lauro, C. (2005), "PLS path modeling", *Computational Statistics and Data Analysis*, Vol. 48 No. 1, pp. 159-205.

van Smeden, M., Moons, K.G., de Groot, J.A., Collins, G.S., Altman, D.G., Eijkemans, M.J. and Reitsma, J. B. (2019), "Sample size for binary logistic prediction models: beyond events per variable criteria", *Statistical Methods in Medical Research*, Vol. 28 No. 8, pp. 2455-2474.

Voss, K.E. (2023), "Composite-based and covariance-based structural equations modeling: moving forward by changing the dialogue", *European Journal of Marketing*, Vol. 57 No. 6, pp. 1780-1792.

Wainer, H. (1976), "Estimating coefficients in linear models: It don't make no nevermind", *Psychological Bulletin*, Vol. 83 No. 2, pp. 213-217.

Widaman, K.F. and Revelle, W. (2023a), "Thinking about sum scores yet again, maybe the last time, we don't know, Oh no . . .: a comment on", *Educational and Psychological Measurement*.

Widaman, K.F. and Revelle, W. (2023b), "Thinking thrice about sum scores, and then some more about measurement and analysis", *Behavior Research Methods*, Vol. 55 No. 2, pp. 788-806.

Wilks, S.S. (1938), "Weighting systems for linear functions of correlated variables when there is no dependent variable", *Psychometrika*, Vol. 3 No. 1, pp. 23-40.

Wold, H. (1982), "Soft modeling: the basic design and some extensions", in Jöreskog, K.G. and Wold, H. (Eds), *Systems under Indirect Observations: Part II*, North-Holland, Amsterdam, pp. 1-54.

Yuan, K.-H., Wen, Y. and Tang, J. (2020), "Regression analysis with latent variables by partial least squares and four other composite scores: consistency, bias and correction", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 27 No. 3, pp. 333-350.

Appendix

To further illustrate the problems that arise from CEI's concealing of measurement model issues, we draw on the model and data from Schlägel and Sarstedt (2016), who researched the impact of four cultural intelligence dimensions (*Cognitive*, *Metacognitive*, *Motivational* and *Behavioral*) on expatriate intentions across various countries. Their research illustrates the challenges of establishing measurement invariance in cross-cultural research. Our analysis focuses on the data from Turkey and USA whose estimates induced adjustments in the original measurement model setups in Schlägel and Sarstedt (2016). Computing the CEI for the Turkey sample produces values between 0.970 (*Metacognitive*) and 1.00 (*Expatriation Intention*), which, according to CEI recommendations, suggests that researchers should use equal weights. However, analyzing this sample with PLS-SEM [Figure A1(a)] identifies a series of indicators with low loadings, particularly in the *Cognitive* construct, which decreases the AVE to 0.459 (i.e. lower than the suggested threshold). Similarly, the low loading of *meta1* pushes the corresponding construct's Rho_A value (0.690) slightly below the suggested threshold. Compared to the equal weights estimation [Figure A1(b)], the PLS-SEM analysis produces a higher R^2 value but lower path coefficient estimates. For example, the effect of *Metacognitive* on *Expatriation Intention* is considerably stronger when using equal weights (-0.122) compared to PLS-SEM (-0.046).

Analyzing the US sample yields CEI values between 0.985 (Cognitive) and 1.000 (*Expatriation Intention*), which would again call for equal weights estimation. Doing so, however, whitewashes convergent validity issues in the *Cognitive* construct (AVE = 0.396), leading to a lower R^2 in the equal weights estimation and clear differences in the path coefficient estimates compared to the PLS-SEM results [Figure A2(a, b)].
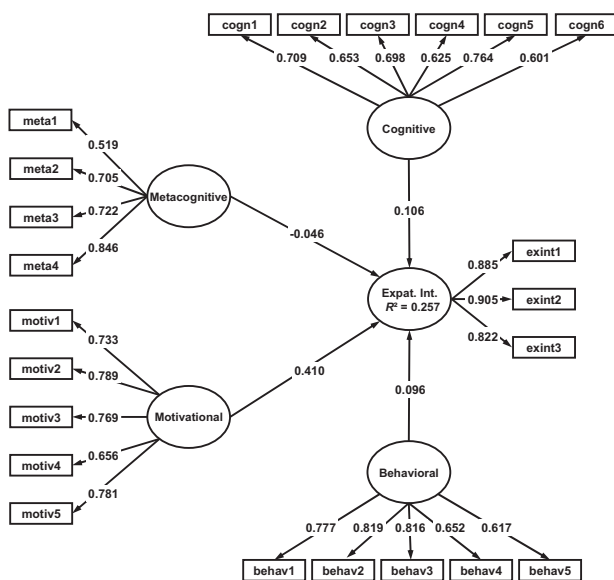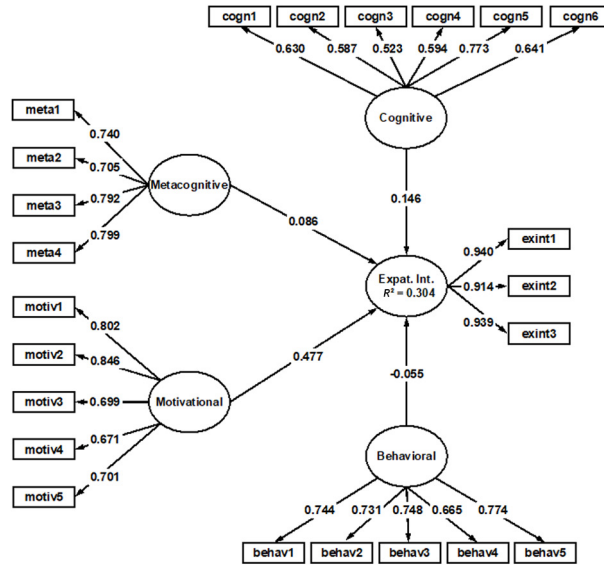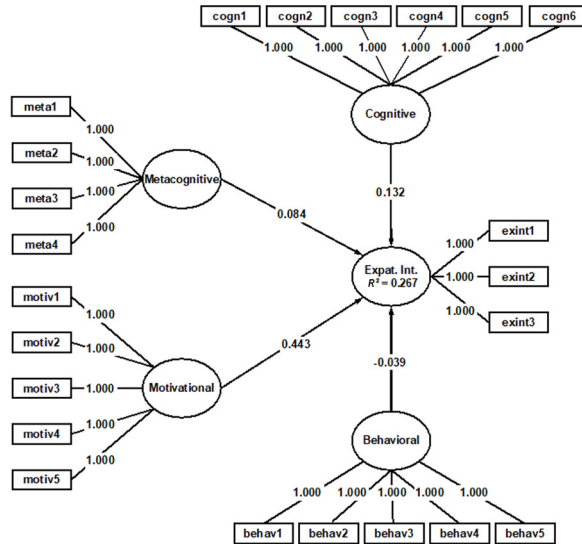
(a)
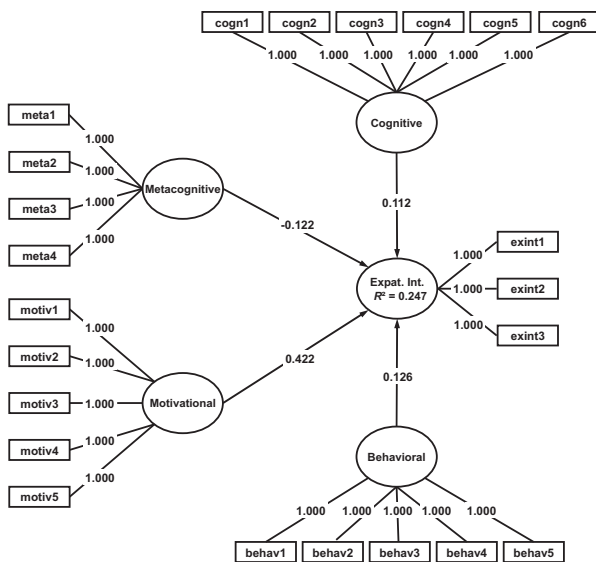
(b)

Notes: (a) PLS-SEM results; (b) equal weights results
Source: Authors' own work

(a)

(b)

Notes: (a) PLS-SEM results; (b) equal weights results
Source: Authors' own work

About the authors

Joseph F. Hair, Jr is Director of the PhD Program and Cleverdon Chair of Business, Mitchell College of Business, University of South Alabama. In 2018, 2019 and 2020, Joe was recognized by Clarivate Analytics for being in the top 1% globally of all Business and Economics professors based on his citations and scholarly accomplishments. He has authored over 75 book editions and has published numerous articles in scholarly journals such as the *Journal of Marketing Research, Journal of Academy of Marketing Science, European Journal of Marketing, Organizational Research Methods, Journal of Family Business Studies, Journal of Retailing*, and others.

Pratyush N. Sharma is an Associate Professor of MIS in the Department of Information Systems, Statistics, and Management Science in the University of Alabama's Culverhouse College of Business. His research interests include online collaboration communities, open-source software development, technology use and adoption and research methods used in information systems, particularly partial least squares path modeling. His research has been published in highly acclaimed journals such as the *Journal of the Association for Information Systems, Journal of Retailing, Decision Sciences, Journal of Information Systems, Journal of Business Research and Journal of International Marketing*.

Marko Sarstedt is a Chaired Professor of Marketing at Ludwig Maximilians University Munich, Germany, and an adjunct research professor at Babeș-Bolyai University, Romania. His main research interest is the advancement of research methods to further the understanding of consumer behavior. His research has been published in *Nature Human Behavior, Journal of Marketing Research, Journal of the Academy of Marketing Science, Multivariate Behavioral Research, Organizational Research Methods, MIS Quarterly, Decision Sciences and Psychometrika*, among others. Marko has been named a member of Clarivate Analytics' Highly Cited Researchers List. In March 2022, he was awarded an honorary doctorate from Babeș-Bolyai-University Cluj-Napoca for his research achievements and contributions to international exchange. Marko Sarstedt is the corresponding author and can be contacted at: sarstedt@lmu.de

Christian M. Ringle is a Chaired Professor of management and decision sciences at the Hamburg University of Technology (TUHH), Germany. His research focuses on management and marketing topics, method development, business analytics, machine learning and the application of business research methods to decision-making. His contributions have been published in journals such as the *European Journal of Marketing, International Journal of Research in Marketing, Information Systems Research, Journal of Business Research, Journal of the Academy of Marketing Science, Long Range Planning, Organizational Research Methods* and *MIS Quarterly*. Since 2018, Dr Ringle has been included in the Clarivate Analytics' Highly Cited Researchers list. He is also a codeveloper and cofounder of the statistical software SmartPLS (www.smartpls.com).

Benjamin D. Liengaard is an Associate Professor in the Department of Economics and Business Economics, Aarhus University. His main research interest is in partial least squares path modeling and quantitative analysis in the field of business analytics. His research has been published in journals such as *Journal of Applied Econometrics, Psychology and Marketing, European Journal of Marketing and Decision Sciences*.