Journal of Business Research 109 (2020) 101–110

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Assessing measurement model quality in PLS-SEM using confirmatory composite analysis

Joe F. Hair Jr.[a,*], Matt C. Howard[a], Christian Nitzl[b]

[a] Department of Marketing & Quantitative Methods, Mitchell College of Business, University of South Alabama, Mobile, AL, USA
[b] Center for Intelligence and Security Studies, University of the German Armed Forces, Munich, Germany

ARTICLE INFO

Keywords:
Confirmatory composite analysis
CCA
PLS-SEM
Confirmatory factor analysis
CFA
Measurement model confirmation

ABSTRACT

Confirmatory factor analysis (CFA) has historically been used to develop and improve reflectively measured constructs based on the domain sampling model. Compared to CFA, confirmatory composite analysis (CCA) is a recently proposed alternative approach applied to confirm measurement models when using partial least squares structural equation modeling (PLS-SEM). CCA is a series of steps executed with PLS-SEM to confirm both reflective and formative measurement models of established measures that are being updated or adapted to a different context. CCA is also useful for developing new measures. Finally, CCA offers several advantages over other approaches for confirming measurement models consisting of linear composites.

1. Introduction

Social scientists have been examining measurement quality for decades. The process of examining and assessing measurement quality is generally associated with the field of psychometrics. This process was initially referred to as item analysis and involved applying statistical methods to select items to include in a psychological test. Guilford (1936) is often identified as the first scholar to refer to the concept of item analysis. The item analysis process varies depending on the context or discipline, but the purpose of item analysis is to develop a relatively small number of items (indicators) that can be used to accurately measure a concept (Crocker & Algina, 1986).

The purpose of this paper is to provide an initial overview of measurement quality assessment, and then to focus on describing the method of confirming measurement quality when applying partial least squares structural equation modeling (PLS-SEM). Specifically, we discuss the emerging process of confirmatory composite analysis (CCA), the measurement model assessment steps in PLS-SEM, compare these steps to confirmatory factor analysis (CFA), and then describe the steps to apply the method, including rules of thumb to guide the researcher in interpreting each stage of the analysis for both reflective and formative measurement models. This research focuses on CCA as the initial step in PLS-SEM since this method is the most popular composite method for estimating SEMs. It should be noted, however, that CCA can be conducted using other composite-based methods, such as generalized structured component analysis (Hwang & Takane, 2004). While our recommendations are positioned in the context of PLS-SEM due to its statistical benefits, the current article and CCA provide implications for the broader scope of composite-based methods (Henseler et al., 2014; Schuberth et al., 2018).

2. Background

The assessment of measurement models was an outgrowth of classical test theory. One of the earliest concepts of classical test theory was proposed by Charles Spearman (1904). He developed a method for correcting a correlation coefficient for measurement error, referred to as attenuation, as well as a reliability index used in making the correction. Other early social scientists contributing to the development of classical test theory include the work by George Yule (1907, 1911), Fritz Kuder (Hakel, 2000), Marion Richardson (Lorr & Heiser, 1965), and Louis Guttman (Stouffer et al., 1950). More recently, in psychometrics and related social sciences classical test theory has been superseded by more sophisticated models such as item response theory (IRT) (Hambleton, Swaminathan, & Rogers, 1991), and generalizability theory (G-theory) (Brennan, 2001).

One method for developing and assessing the quality of measurement – exploratory factor analysis (EFA) – was initially proposed by Charles Spearman (1904). Researchers have often applied EFA for data reduction and exploration of the theoretical structure of psychological phenomena. The statistical objective of EFA is to identify a set of latent constructs from a large number of individual variables (items), with the result being reliable and valid measurement scales. The method is

* Corresponding author.
E-mail addresses: jhair@southalabama.edu (J.F. Hair), mhoward@southalabama.edu (M.C. Howard), christian.nitzl@unibw.de (C. Nitzl).

https://doi.org/10.1016/j.jbusres.2019.11.069
Received 9 August 2019; Received in revised form 20 November 2019; Accepted 22 November 2019
Available online 10 December 2019

typically applied when there are no prior hypotheses about factors or patterns of measured variables. Researchers start with a large number of observed variables assumed to be related to a smaller number of 'unobserved' factors and expect to reduce the large number of variables to a much smaller set of factors (Brown, 2014; Kline, 2014, 2015). For example, a researcher may develop a customer satisfaction construct consisting of ten items, and EFA could be used to determine whether a single factor adequately explains the variance and covariance among these ten items. The EFA could also determine whether each item sufficiently relates to the emergent factor, or whether certain items should be removed because they are not representative of the proposed factor. Examples of constructs derived from measured variables in the social and behavioral sciences could be trust, satisfaction, commitment, coordination, performance, loyalty, and so forth. As a result, exploratory factor analysis (EFA) has become an essential tool for research in the social sciences, business, and beyond (Grimm & Yarnold, 1995; Hair, Black, Anderson, & Babin, 2019; Howard, 2016; Kline, 2015; Rencher & Christensen, 2012).

Early applications of EFA were based on the common factor model. In the common factor model, measured variables are assumed to be a function of common variance, specific variance, and error variance (Brown, 2014; Hair, Black, et al., 2019). Common variance is the indicator's variance that is shared with all other indicators in the analysis. Specific variance is the indicator's variance that is only associated with the indicator. Error variance is the indicator's variance that is due to unreliability, bias, or randomness. EFA assumes that any indicator/measured variable may be associated with any factor. When developing a measurement scale, researchers may use EFA first to examine the underlying structure of multi-item scales (Brown, 2014; Hair, Black, et al., 2019; Kline, 2014; Kline, 2015) before moving on to CFA. However, EFA is not a required step in scale development, and only needs to be applied when there is no established theory describing the underlying factors/constructs for a set of measured variables.

EFA requires the researcher to make a number of important decisions about how to execute the analysis. For example, how many factors to extract to represent the underlying patterns, the size of the loadings, and which type of rotation to apply? Perhaps most important, however, is choosing which of two primary approaches should be applied, since each of the approaches is applied in a different research context. One approach is principal axis analysis, which extracts factors using only common (shared) variance. The other approach is principal components analysis (PCA), which extracts factors using total variance.

Common factor analysis uses only common variance and assumes specific and error variance is unimportant in defining the factors (Brown, 2014; Kline, 2014; Thompson, 2004). On the other hand, principal components (composite) analysis is focused on explaining as much of the total variance as possible, and results in factors that contain primarily common variance, and also some specific and error variance (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Preacher & MacCallum, 2003). Methodologists advocating PCA note that the amount of error variance included when starting with total variance is most often negligible (Bentler & Kano, 1990; Hair, Black, et al., 2019) because the process of extracting factors removes most if not all of the error variance. Some methodologists have argued that component (composite) analysis is always more appropriate than common factor analysis while others believe common factor analysis is always more appropriate than component analysis. Based on our review it is most likely neither side is entirely correct and both approaches are appropriate for specific research applications (Brown, 2014; Grimm & Yarnold, 1995; Hair, Black, et al., 2019; Howard, 2016; Kline, 2014; Kline, 2015; Thompson, 2004). It should also be noted that PCA and principal axis analysis converge in many situations commonly encountered in applied research. The two types of approaches provide very similar results when studying more than 30 indicators and/or communalities exceed 0.60, whereas their differences are more pronounced when studying less than 20 indicators and/or communalities are below 0.40 (Stevens, 2009).[1] Thus, the choice of analytical approach may have a small influence in many research scenarios.

More recently the concepts of exploratory factor/component analysis have been extended and applied to confirmation of theoretical structures, particularly measurement models for psychological constructs/concepts. The emergence of covariance-based structural equation modeling (CB-SEM) in the early 1980's included a process for assessing measurement model quality referred to as confirmatory factor analysis (CFA). CFA is both a qualitative and statistical process that involves examining the reliability of the individual indicators (item reliability), construct reliability, qualitative face and content validity, quantitative measures of convergent and discriminant validity, and Goodness of Fit.

The CFA process enabled researchers to evaluate multi-item constructs and scholars began referring to measurement model confirmation as CFA because constructs were based on common variance and referred to as factors – hence CFA. When researchers apply CFA, established – or at minimum some – meaningful theory is available about the relationships between the individual variables and how they relate to theoretical concepts. In short, researchers are testing the hypothesis that a proposed theoretical relationship exists between the observed variables and their underlying latent construct(s). Moreover, the objective is to confirm the measurement properties of a set of variables (indicators) for measuring a specified and operationally defined latent construct.

PLS-SEM emerged at the same time as CB-SEM. PLS-SEM, also referred to as variance based SEM, was developed by Herman Wold (1982) to offer a structural equation modeling approach with much greater flexibility compared to CB-SEM. The evolution of both approaches was not parallel, however, mainly because of the lack of software to execute PLS-SEM (Mateos-Aparicio, 2011). When applications of PLS-SEM began increasing around 2005, with the availability of PLS Graph (Chin, 2003) and the SmartPLS 2 (Ringle, Wende, & Will, 2005) software, and more recently ADANCO (Henseler & Dijkstra, 2015) and SmartPLS 3 (Ringle, Wende, & Becker, 2014) methodologists initially referred to measurement model confirmation using PLS-SEM as a CFA, just as with CB-SEM. In those early years of PLS-SEM applications, confusion arose around referring to the process of confirming measurement models for both CB-SEM and PLS-SEM as a CFA. As a result, some scholars suggested that PLS methodologists adopt separate terminology for PLS-SEM. In response to this, Ed Rigdon (2014) suggested that PLS-SEM applications and terminology could be clarified if measurement models/constructs are identified as "composites", and that terminology began to be adopted by PLS methodologists (Sarstedt, Ringle, Henseler, & Hair, 2014). That same year Henseler et al. (2014) proposed the concept of CCA as a process for confirming measurement models in PLS-SEM. Next, Sarstedt, Hair, Ringle, Thiele, and Gudergan (2016) published an article that referred extensively to the term composites, even going so far as to suggest that CB-SEM constructs and PLS-SEM constructs are both composites that estimate proxies for theoretical concepts/constructs.

The methods of EFA, CCA and CFA have similarities but many differences. For example, the statistical objective of EFA is data reduction through exploration of response patterns, while the statistical objective of CCA and CFA is confirmation of measurement theory. Other primary differences include the execution of EFA often ends with the identification of factors, while CCA and CFA begin with proposing theoretical constructs to be confirmed, and almost always moves on to structural modeling after the composite measurement models have been confirmed. Further differences among the three methods are compared in Table 1.

Before executing a CCA, operational definitions of the multi-item construct must be confirmed, including whether the appropriate measurement model is reflective or formative, since the process for the two

[1] We thank an anonymous reviewer for raising this point.

Table 1
Comparison of EFA, CCA and CFA.

EFA Exploratory Factor Analysis	CCA Confirmatory Composite Analysis	CFA Confirmatory Factor Analysis
• Principal Components Analysis (PCA) = Total Variance • Common Factor Analysis = Common Variance	Total Variance	Common Variance Only
Exploratory Only	Both Exploratory and Confirmatory	Confirmatory Only
Analyzes Independent and Dependent Variables Separately	Analyzes Independent and Dependent Variables Together, but Focuses on Measurement Confirmation	Analyzes All Variables Together as Measurement Models
Objective is Data Reduction	Objective is Confirming Measurement Models and also Prediction of Dependent Variables	Objective is Confirming Measurement Models
Orthogonal Rotation produces Independent (uncorrelated) Factors	Composites (constructs) are Correlated	Composites (constructs) are Correlated
Reliability Examined Typically Cronbach's Alpha	Reliability Examined Typically Composite Reliability Reflective Measurement Models	Reliability Examined Typically Composite Reliability Reflective Measurement Models
Face and Content Validity	Convergent Validity Reflective Measurement Models	Convergent Validity Reflective Measurement Models
Other Types of Validity typically not Assessed	Discriminant Validity	Discriminant Validity
Factor Scores and Sum Scores often used with Multiple Regression	Construct Composite Scores applied in Structural Modeling	Construct Latent Factors applied in Structural Modeling
Factor Scores are Indeterminant	Construct Composite Scores are Determinant	Construct Factor Scores are Indeterminant

types of measurement differs considerably. When the measurement theory is less developed, these initial steps are followed by literature reviews and qualitative research with expert panels to assess face validity and reduce the initial list of items (Goetz et al., 2013; Howard, 2018). Pilot testing for refinement and purification of the items prepares the researcher for executing a CCA.

CCA differs from CFA in that the statistical objective is to maximize variance extracted from the exogenous variables, but in doing so to facilitate prediction of the endogenous constructs, and confirmation of the measurement models. That is, CCA enables researchers to develop and validate measures within a nomological network. Each composite, therefore, must relate to at least one other composite. Hence, the validity of a composite depends on the nomological network in which it is embedded. The method is an extension of principal components analysis because it is composite-based, and therefore produces composite scores that are weighted linear combinations of indicators and can be used in follow-up analyses. The resulting composites are correlated, as they would be in an oblique rotation with an EFA and include variance that maximizes prediction of the endogenous constructs. Note that the composite correlations from the oblique rotation do not often result in problems with multicollinearity, but this issue should always be examined (Hair, Hult, Ringle, & Sarstedt, 2017).

To achieve measurement confirmation objectives in developing or adapting multi-item measures, researchers could use either CFA or CCA. The results are different, however, and researchers need to understand the implications of the distinct outcomes to make informed decisions. CCA and CFA can both be used to improve item and scale reliability, identify and provide an indication of items that need to be revised or in some instances eliminated for content validity, facilitate achieving convergent validity and discriminant validity, and to remove error variance. Compared to CFA, CCA has several benefits, as follows: (1) the number of items retained to measure constructs is higher with CCA, thereby improving content coverage and construct validity, (2) determinant construct scores are available (Rigdon, Becker, & Sarstedt, 2019), and (3) CCA can be applied to formative measurement models.

Further, one of the most important differences between a CCA and a CFA is the application of so-called goodness-of-fit (GOF) indices. The evaluation of goodness-of-fit is an essential component in completing a CFA. GOF indices evaluate the difference between the variance–covariance matrix using an empirical sample and the estimated model variance–covariance matrix based on the modeled construct measurement (Benitez, Henseler, Castillo, & Schuberth, 2019). The discrepancy between two matrices is measured by criteria like the squared Euclidean distance (d_{ULS}), the geodesic distance (d_G), and the standardized root mean squared residual (SRMR). A requirement of CFA solutions is to minimize the difference between the empirical (observed) and the estimated variance-covariance matrix, and minimum guidelines must be achieved to move on to test the structural model.

One stream of research has suggested GOF indices, which are typically used in CFA, are also applicable in a CCA (Dijkstra & Henseler, 2015; Henseler et al., 2014; Schuberth, Henseler, & Dijkstra, 2018). Schuberth et al. (2018) in a simulation study attempted to illustrate the performance of bootstrap-based tests and discrepancy measures as one approach to evaluate overall fit for CCA. However, instead of using a PLS-SEM algorithm to construct a composite they used a generalized canonical correlation analysis (GCCA) with maxvar option. Furthermore, the authors used normally distributed datasets in their simulations whereas PLS-SEM does not assume normally distributed data. Above that, Schuberth et al. (2018) stated that caution is also needed in case of small sample sizes where misspecified models with the tested GOF indices were not reliably detected. But most studies using PLS-SEM are based on much smaller data sets (Hair, Sarstedt, Ringle, & Mena, 2012; Nitzl, 2016). These technical aspects make the transferability of the results of the simulation regarding GOF indices highly questionable in a PLS-SEM context.

There are also critical conceptual issues that make the transferability of GOF indices to a CCA in PLS-SEM questionable. PLS-SEM is based on causal-predictive relations because it maximizes the amount of explained variance of dependent variables founded in well-developed explanations (Hair, Sarstedt, & Ringle, 2019; Jöreskog & Wold, 1982). Consequently, an estimated composite always depends on the nomological network in PLS-SEM.[2] This is true for both formative and reflective measurement models. In contrast, CFA facilitates a stand-alone evaluation of factor measurements and it is not necessary for it to be embedded in a nomological network. The need to validate composites in a nomological network using PLS-SEM means the same composite might fit well in one model but not in the other. In a situation where GOF indices indicate a misspecification, the question of whether a researcher should change the indicators of the composite or the nomological network is uncertain and debatable, and if indicators are changed the results are likely to compromise validity and reliability metrics.

Additionally, using GOF indices in CCA means that a researcher

[2] This is also true in the application of GOF for the CCA approach described by Schuberth et al. (2018).

focus only on the explanation characteristics of PLS-SEM and ignore the predictive characteristics in evaluating the model results (Evermann & Tate, 2016; Hair, Sarstedt, et al., 2019; Shmueli, Ray, Velasquez Estrada, & Chatla, 2016). Testing GOF for PLS-SEM is only useful for studies that follow a purely confirmatory approach, whereas PLS-SEM requires consideration of confirmation, explanation, and prediction. The question arises why such a benchmark for evaluation based on a purely confirmatory viewpoint should be used when it is not the main objective of the analysis? Application of GOF indices in such a context is questionable in the best case, and in the worst case applying GOF indices could even reduce the predictive power of a CCA (Hair, Hult, et al., 2017; Lohmöller, 1989; Rigdon, 2012). The logical and more promising approach should rely on criteria that emphasize the confirmation, explanation and predictive characteristics of PLS-SEM. One possibility would be information theoretic model selection criteria that build on a tradeoff between model fit and predictive power (Sharma, Sarstedt, Shmueli, Kim, & Thiele, 2019; Sharma, Shmueli, Sarstedt, Danks, & Ray, 2019). Their use is restricted thus far, however, to model comparisons.

The methods of EFA and CFA are well established in the literature (Hair, Risher, Sarstedt, & Ringle, 2019). In contrast, CCA as a separate approach to confirming linear composite constructs in measurement models is just emerging (Hair, Black, et al., 2019; Hair, Page, & Brunsveld, 2020; Henseler, Hubona, & Pauline, 2016; Schuberth et al., 2018). The remaining focus of this article will therefore be on explaining the steps followed to apply CCA.

3. Confirmatory composite analysis

In recent years researchers have begun referring to the measurement model assessment step in PLS-SEM as CCA (Henseler et al., 2014; Schuberth et al., 2018). CCA is a systematic methodological process for confirming measurement models in PLS-SEM.

3.1. Reflective measurement models

When conducting a CCA, it is important for researchers to understand that reflective measurement models are composite latent constructs whose indicators (measured variables) are assumed to be influenced, affected, or caused by the underlying latent variable (Sarstedt et al., 2016). A change in the latent construct will be reflected in a change in all of its indicators. The indicators are seen as a manifestation of the empirical surrogates (proxy variables) for the latent variable. In contrast, the indicators of a formative composite latent construct are viewed as causing rather than being caused by the underlying latent construct. With formative measurement models, a change in the latent construct is not necessarily accompanied by a change in all of the indicators. But if any one of the indicators is removed, or if a new indicator is added, then the definition of the latent construct would change.

A researcher should be aware the above description of reflective and formative measurements refers to the epistemic relationship between indicators and constructs as assumed from measurement theory (Sarstedt et al., 2016). PLS-SEM, including the CCA approach, computes composites from linear combinations of sets of indicators to represent the concepts in the statistical model. If correlation weights are used for estimating a composite, the arrows typically point away from a construct to the indicators. This is often referred to as a reflective measurement model. However, if regression weights are used for compositing a composite, the arrows typically point from the indicators to their construct. This is often referred to as a formative measurement (Nitzl & Chin, 2017).

The steps listed in Table 2 should be followed to execute a CCA with reflective measurement models. These steps are described in detail below:

Step 1: Assessing the indicator loadings and their significance. The standardized loadings should have a value of at least 0.708 and an associated t-statistic above ± 1.96 to be significant for a two-tailed test at the 5% level (Hair, Ringle, & Sarstedt, 2011). T-statistics in PLS-SEM are obtained by executing a bootstrapping procedure (Hair, Sarstedt, et al., 2012). Alternatively, Wood (2005) introduced the use of confidence intervals with PLS-SEM. Indicator loadings confidence intervals can be used in a manner similar to t-statistics and intervals excluding zero are statistically significant. A benefit of confidence intervals is the dichotomous approach of significance testing is avoided and authors are able to consider other methods to identify practically significant indicator loadings when using confidence intervals (Cohen, 1994).

Table 2
Steps in Confirmatory Composite Analysis with Reflective Measurement Models.

Assessing Reflective Measurement Models Using Confirmatory Composite Analysis (CCA)
1. Estimate of Loadings and Significance
2. Indicator Reliability (items)
3. Composite Reliability (construct)
4. Average Variance Extracted (AVE)
5. Discriminant Validity – HTMT
6. Nomological Validity
7. Predictive Validity

Step 2: Squaring the individual indicator loadings provides a measure of the amount of variance shared between the individual indicator variable and its associated construct. This is referred to as indicator reliability (Hair, Black, et al., 2019).

Step 3: The reliability of the construct can be measured in two ways – Cronbach's alpha (α) and composite reliability (CR). The rule of thumb for both reliability criteria is they need to be above 0.70. Because indicators are not equally reliable, composite reliability, which is weighted, is more accurate then Cronbach alpha (unweighted), and therefore CR should be assessed and reported (Hair et al., 2019). Note that internal consistency reliability, including both Cronbach alpha and composite reliability, can be too high. If reliability is 0.95 or higher, the individual items are measuring the same concept, and are therefore redundant. In short, redundancy indicates the indicators are measuring the same concept and therefore do not include the required diversity to ensure the validity of multi-item constructs (Hair, Risher, et al., 2019).

Step 4: Convergent validity can be measured by the Average Variance Extracted (AVE). The AVE is obtained by averaging the indicator reliabilities of a construct. This metric measures the average variance shared between the construct and its individual indicators. The criterion for AVE is the value should be 0.5 (50%) or higher.

Step 5: Discriminant validity measures the distinctiveness of a construct. Discriminant validity is demonstrated when the shared variance within a construct (AVE) exceeds the shared variance between the constructs. The method that should be used is the heterotrait-monotrait ratio of correlations (HTMT) (Henseler, Ringle, & Sarstedt, 2015). Researchers can apply cutoff scores such as 0.85 and 0.90 to interpret their HTMT results. Additionally, Franke and Sarstedt (2019) recently recommended additional significance testing that includes confidence intervals to further assess HTMT ratios and discriminant validity.

Step 6: Nomological validity is an additional method of assessing construct validity. The process for measuring nomological validity is to correlate the construct score of each construct with one or more other constructs (concepts) in the nomological network. The nomological network (or nomological net) is a representation of the concepts (constructs) that are the focus of a study as well as the interrelationships between the concepts (Cronbach & Meehl, 1955). The other constructs represent key relationships and are often classification variables such as employee age, years of experience, part-time versus full-time workers, and so forth. The other constructs may also be concepts not included in the theoretical model being tested. Previous research results are used to identify whether the theoretical relationship with the other constructs is positive or negative, as theory would suggest.

To further clarify nomological validity, assume you are examining a structural model consisting of an exogenous variable job satisfaction and an endogenous construct job performance. For this theoretical model, data is also collected on organizational commitment and job type (full-time versus part-time), constructs that are in the nomological net of job satisfaction and job performance. Nomological validity can be assessed by correlating the construct scores of job satisfaction and job performance with organizational commitment and job type to determine if the results are consistent with the theoretical direction as well as the size and significance of the correlations.

Step 7: Predictive validity assesses the extent to which a construct score predicts scores on some criterion measure. Predictive validity is similar to concurrent validity since both types are measured in terms of the correlation between a construct score and some other criterion measure. The difference is that predictive validity involves using the construct score to predict the score of a criterion variable that is collected at a later point in time, while concurrent validity assesses the correlation between the scores of two variables when the data is collected at the same time. For example, if job satisfaction and organizational commitment data are collected at the same time, and the scores are correlated and the results are consistent with theory (direction and significance are as expected), then concurrent validity is established. Using the same constructs, if data on job satisfaction is collected first, and organizational commitment data is collected six months later, then the two construct scores could be correlated, and the result would assess predictive validity.

Although not technically a method of predictive validity, construct invariance in PLS-SEM measurement models can be tested by applying the MICOM procedure (Henseler, Ringle, & Sarstedt, 2016). Construct invariance (equivalence) is most often applied with cross-cultural studies and can be assessed with MICOM. But application of invariance by comparing measures over time is also possible. Applying MICOM to longitudinal and/or casual effects ensures that observed changes, if any, are due to substantive relationships of the constructs rather than changes in the nature of the constructs themselves (Vandenberg & Lance, 2000). Assessment of longitudinal measurement invariance is not a required portion of CCA or other PLS-SEM analyses (Hair, Ringle, & Sarstedt, 2012; Hair, Sarstedt, Ringle, & Gudergan, 2018), unless cross-cultural studies are a characteristic of the research design.

3.2. Formative measurement models

When executing CCA with formative composite measurement models, researchers must remember that formative measurement models differ from reflective measurement models. Formative composite measurement models are linear combinations of a set of indicators that form the construct. That is, the indicators point from the measured variables to the composite construct, are considered causal, and do not necessarily covary. As a result, the underlying internal consistency concepts associated with reflective measurement models cannot be applied to formative measurement models. Formative indicators cannot be evaluated using composite reliability or AVE (Chin, 1998), and indicator loadings in formative models must be interpreted in a different manner than those in reflective models (described below). This characteristic is particularly applicable to PLS-SEM because the method assumes formative composite indicators completely capture the entire domain of the construct under consideration (Hair, Hult, et al., 2017; Sarstedt et al., 2016).

As noted above, composite reliability and AVE are not appropriate metrics to evaluate formative measurement models. In addition, as with the evaluation of reflective measurement models, goodness of fit measures are not required when executing CCA with formative measurement models (Hair, Hult, et al., 2017; Hair, Matthews, Matthews, & Sarstedt, 2017; Hair, Sarstedt, et al., 2019). The steps listed in Table 3 should be followed to execute a CCA with formative measurement models.

Table 3
Steps in Confirmatory Composite Analysis with Formative Measurement Models.

Assessing Formative Measurement Models Using Confirmatory Composite Analysis (CCA)
1. Convergent Validity – redundancy
2. Indicator Multicollinearity
3. Size and Significance of Indicator Weights
4. Contribution of Indicators (size & significance of loadings)
5. Assess Predictive Validity

Step 1: Convergent validity with formative measurement models is the extent to which the formative construct is positively correlated with a reflective measure(s) of the same construct using different indicators (Hair, Hult, et al., 2017). The relationship between the multi-item formative construct and the reflective measure of the same construct is typically examined using correlation or regression. The reflective measure is often a single item, but as in other situations when measuring ambiguous concepts multi-item reflective measures are preferred. Cheah, Sarstedt, Ringle, Ramayah, and Ting (2018) note when the sample size is small a single item measure exhibits higher degrees of convergent validity than a reflective multi-item construct does. But for larger sample sizes the differences are marginal. Hence, the use of a global, reflectively measured single item in PLS-SEM-based redundancy analyses is sufficient.

When the research involves survey data, the reflective measure for testing convergent validity must be included in the research design as part of the data collection process. If the theoretical SEM is based on secondary data, the researcher should identify another secondary measure of the same construct, which could be either reflective or formatively measured, to use as a proxy variable for testing the convergent validity of formative constructs. Acceptable endogenous reflectively measured constructs to use as proxy variables in testing convergent validity can be identified by reviewing established scales from previously published research. Another option is to develop a more general global item that summarizes the essential concepts of the formative construct (Hair, Hult, et al., 2017).

Demonstration of convergent validity is based on the size of the path coefficient between the two constructs. The formative construct is the exogenous variable and the reflective measure is the endogenous variable. As a guideline, we recommend a minimum path coefficient of 0.70, and the larger the size of the coefficient the stronger the indication of convergent validity (Hair, Hult, et al., 2017). If convergent validity is unable to be established using this criterion, the researcher should determine if revision of the definition of the theoretical formative construct is possible by removing, revising, or adding one or more indicators. The process of examining convergent validity with formative measurement models is also referred to as redundancy analysis (Chin, 1998).

Step 2: Indicator multicollinearity is the extent to which the formative items are correlated. Recall that reflective indicators are often considered to be interchangeable and thus high correlations are anticipated. But high correlations between formative indicators create problems with multicollinearity. Calculation of formative construct scores is based on a multiple regression model, with the dependent variable being the construct score and the formative indicators the independent variables. Just as high multicollinearity creates problems with multiple regression models, it also creates problems with formative measurement models, such as distorting the size of the beta coefficients (weights) and/or changing the sign of these same coefficients.

To determine if multicollinearity is a problem, researchers must assess whether it is present. To do so, the VIF (variance inflation factor) available in most statistical software is examined. If the VIF is 3.0 or lower, then multicollinearity is unlikely to be a problem. Note that in previous publications the acceptable VIF level was thought to be 5.0, but subsequent research indicates this level is too high (Hair, Black, et al., 2019; Hair et al., 2020). An alternative approach to examining the level of multicollinearity is to calculate bivariate correlations between the formative indicators. If the bivariate correlations between formative indicators are 0.50 or higher that is an indication the levels of multicollinearity are too high and will create problems.

When multicollinearity is high among formative indicators, the researcher can evaluate whether one or more indicators can be removed. But recall that removing formative indicators often changes the operational definition of the construct, so researchers must consider this possibility as well. Another option to resolve high multicollinearity among formative indicators is to develop higher-order constructs (HOCs) that are supported by measurement theory (Becker, Klein, & Wetzels, 2012; Kuppelwieser & Sarstedt, 2014; Ringle, Sarstedt, & Straub, 2012). For additional information on the topic of creating higher-order constructs, see Hair et al. (2018) and Sarstedt, Hair, Cheah, Becker, and Ringle (2019).

Step 3: If the metrics obtained in the previous CCA steps indicate the formative measurement model meets recommended guidelines, the researcher next examines the size and significance of the indicator weights. The purpose of this step is to determine the extent to which the formative indicators contribute to the construct score. The amount of contribution (relevance) of the indicators is interpreted based on the size of the outer model weights, with larger weights indicating a higher contribution. The weights are calculated using a multiple regression model and are equivalent to beta coefficients. They represent the relative contribution of each formative indicator in forming the construct. Therefore, the values of the outer weights are almost always smaller than the outer loadings on reflectively measured constructs.

In addition to the sizes of the outer weights, the values of the formative indicator weights must be statistically significant. PLS-SEM is a nonparametric statistical method, so significance is determined using bootstrapping. In general, the level of statistical significance required is ≤ 0.05. But when PLS models are tested using small sample sizes, it may be justifiable to lower the acceptable level of significance to ≤ 0.10.

An important issue to consider when evaluating formative measurement models is the number of indicators. As the number of indicators increases, the likelihood of one or more indicators having low or even non-significant outer weights also increases. When faced with this situation, the recommended solution is the same as when high multicollinearity is present – to create higher-order constructs by combining the formative indicators into conceptually similar and theoretically supportable lower-order constructs (Cenfetelli & Bassellier, 2009). For more guidance on this approach see Hair, Hult, et al. (2017), and Sarstedt, Hair, et al. (2019). Since creating higher-order constructs is not always possible, an alternative approach is outlined in Step 4 for moving ahead with a formative construct CCA when the outer weights are small and not significant.

Step 4: Assessing the absolute contribution of formative indicators can also be applied to justify retaining formative indicators. The absolute contribution of a formative indicator is the amount of information contributed by the indicator in forming the construct, if no other indicators are considered in the calculation. The absolute contribution is derived from the formative indicator's outer loading. The outer loading is equivalent to the bivariate correlation between each indicator separately and the construct. The loading is considered absolutely important in forming the formative construct when it is ≥ 0.50 and statistically significant. But the outer loading is not an indication of importance. As a final note, in conducting a CCA researchers occasionally encounter a situation with a formative indicator whose absolute contribution is lower than < 0.50 and not significant. In such situations, the researcher can remove or retain the formative indicator based on a theoretical assessment of its relevance obtained from knowledgeable experts (Cenfetelli & Bassellier, 2009). In sum, a formative indicator should never be eliminated based solely on statistical criteria.

Table 4
Steps in Structural Model Assessment.

Assessing Structural Model Results
1. Evaluate structural model collinearity
2. Examine size and Significance of Path Coefficients
3. R^2 of Endogenous Variables (in-sample prediction)
4. f^2 Effect Size (in-sample prediction)
5. Predictive Relevance Q^2 (primarily in-sample prediction)
6. PLSpredict (out-of-sample prediction)

Step 5: As with reflectively measured constructs, predictive validity assesses the extent to which a construct score predicts scores on some criterion measure. As noted earlier, predictive validity involves using the construct score to predict the score of a criterion variable that is collected at a later point in time. For example, if data using formative constructs is collected first, and data on the criterion constructs is collected six months later, then the two construct scores could be correlated, and the result would assess predictive validity.

4. Structural model assessment

The primary focus of this article is to introduce and explain the process of CCA. Major recent developments in PLS-SEM lead us to also include a brief summary of the steps in assessing the structural model, with particular emphasis on the latest methodological improvement – PLSpredict. Recall that before assessing structural model results, the measurement models must first be confirmed using the CCA process. The steps listed in Table 4 should be followed to assess the structural model.

Step 1: Evaluation of the structural model results relies heavily on the concepts and characteristics underlying multiple regression analysis. As a result, the first step is to evaluate the structural model constructs to determine if high multicollinearity is a problem. Structural models characterized by high multicollinearity can affect the size of the beta coefficients (weights) by increasing or decreasing them and/or changing the signs of these same coefficients. As with indicators on formative constructs, the VIF values can be examined and if they are below 3.0, then multicollinearity is unlikely to be a problem. An alternative approach is to examine the bivariate correlations between the construct scores. If the bivariate correlations are higher than 0.50 multicollinearity could influence the size and/or signs of the path coefficients. When multicollinearity appears to be a problem, the recommended solution is to create higher-order constructs by combining the separate constructs into conceptually similar and theoretically supportable lower-order constructs (Cenfetelli & Bassellier, 2009).

Step 2: If multicollinearity is not a problem, the second step is to examine the size and significance of the path coefficients. This process enables the researcher to test the hypothesized relationships among the constructs. The path coefficients are standardized values that may range from $+1$ to -1, but they seldom approach $+1$ or -1. This is particularly true with complex models having multiple independent constructs in the structural model. The closer the path coefficient values are to 0 the weaker they are in predicting dependent (endogenous) constructs, and the closer the values are to the absolute value of 1 the stronger they are in predicting dependent constructs.

Note that as the final step in CCA the researcher should examine the predictive ability of the structural model. Steps 3 through 6 below represent the four metrics to apply in examining structural model prediction.

Step 3: As with multiple regression models, the most often used metric to assess structural model prediction is R^2. Referred to as the coefficient of determination, it is a measure of in-sample prediction of all endogenous constructs. That means the prediction is a measure of the predictive ability only for the sample of data used in calculating the results, and R^2 should not be inferred to the population (Rigdon, 2012; Sarstedt et al., 2014). The minimum R^2 value is 0 but it would almost never be that low. As with multiple regression, the more independent variables (constructs) in the structural model the higher the R^2, assuming the independent variables are in fact related to the dependent variable constructs. The maximum R^2 value is 1 but values this high very seldom occur. In evaluating the size of the structural model R^2, the researcher should review similar research in relevant empirical research and use those results as a guideline, assuming the context of the research is not too different. Finally, some disciplines also examine the adjusted R^2, which systematically adjusts the R^2 value downward based on the sample size and the number predictive constructs. As with multiple regression, the adjusted R^2 is useful when researchers include too many nonsignificant predictor constructs in the structural model (Hair, Hult, et al., 2017).

Step 4: A second measure of the predictive ability of the structural model is the effect size, which provides an estimate of the predictive ability of each independent construct in the model. To calculate this value, each predictor construct is systematically removed from the model (SmartPLS does this automatically) and a new R^2 is calculated without the predictor. Next the R^2 with the predictor in the model is compared to the R^2 without the predictor in the model, and the difference in the two R^2 values indicates whether the omitted construct is a meaningful predictor of the dependent construct (Hair, Hult, et al., 2017). The effect size, referred to as an f^2, is ranked as small, medium and large. Values above 0.02 and up to 0.15 are considered small; values of 0.15 and up to 0.35 are medium; and values 0.35 and above are large effects (Cohen, 1988). The effect size is also considered an in-sample predictive metric.

Step 5: A third metric used to assess prediction is the Q^2 value, also referred to as blindfolding (Geisser, 1974; Stone, 1974). Some scholars consider this metric to be an assessment of out-of-sample predictive power, and to some extent it is. But it clearly is not as strong a model prediction metric as is PLSpredict, described in the next step. When interpreting Q^2, values larger than zero are meaningful whereas values below 0 indicate a lack of predictive relevance. In addition, Q^2 values larger than 0.25 and 0.50 represent medium and large predictive relevance of the PLS-SEM model.

Step 6: For some types of research in-sample prediction is sufficient. But in other situations, researchers need a robust method to assess out-of-sample prediction. The above-mentioned criteria for evaluating predictive validity, including R^2, f^2 and to some extent Q^2, are useful in evaluating the predictive power of a model based on in-sample (Sarstedt et al., 2014). In-sample prediction uses the same sample to estimate the model and to predict responses, which is likely to overstate the model's predictive ability. This is typically referred to as an overfitting problem (a higher prediction than is realistic), and indicates the model may have limited value in predicting observations not in the original sample. Shmueli et al. (2016) recently proposed an approach to assess out-of-sample prediction when using PLS-SEM. The approach involves first estimating the model on an analysis (training) sample and using the results of that model to predict other data in a second separate holdout sample.

The PLSpredict method (Shmueli et al., 2019) generates holdout-based sample predictions in PLS-SEM and is an option in standard PLS-SEM software so researchers can apply the method. For example, SmartPLS (Ringle et al., 2014) and open source software such as R (https://github.com/ISS-Analytics/pls-predict) include PLSpredict. PLSpredict first randomly splits the total sample into subgroups that are equal in size. Each subgroup is called a fold and the number of subgroups is k. If you divide the total sample into 10 groups (folds) then $k = 10$. The method then selects nine subgroups (k-1) and combines them into a single analysis sample. The remaining subgroup becomes the holdout sample that the analysis sample attempts to predict. The process of predicting one of the subgroups rotates through all of the original 10 subgroups so that each of the 10 subgroups is used by itself as a holdout sample, and the other 9 subgroups are combined to predict that holdout sample. When this process is completed, the data for every respondent in the total sample has been predicted by an analysis sample that did not include that respondent's data to estimate the model results.

An important consideration is the size of the holdout sample must be sufficient to produce a robust estimate. If the number of subgroups chosen when running PLSpredict is too large, the size of holdout sample may be too small. For example, if the total sample size is 100, and the number of subgroups is 10, then the holdout sample size would be N = 10, which in most circumstances would be considered too small. Shmueli et al. (2019) recommend setting $k = 10$ subgroups, but if the holdout sample is too small with $k = 10$ then a smaller number of subgroups can be selected. For example, k = 5 subgroups with a total sample of N = 100 would result in a holdout sample size of N = 20.

In general, the recommended minimum size for the holdout sample would be N = 30. But analytical methods such as logistic regression, discriminant analysis, and MANOVA have produced robust results with group sizes as small as N = 20 (Hair, Black, et al., 2019) so there is some flexibility in determining the minimum sample size of the holdout sample subgroup. If the holdout sample size is <30 researchers should interpret the findings cautiously.

For the assessment of a model's predictive power when using PLSpredict, researchers can draw on several prediction statistics that quantify the amount of prediction error. For example, the mean absolute error (MAE) measures the average magnitude of the errors in a set of predictions without considering their direction (over or under). Another popular prediction metric is the root mean squared error (RMSE), which is defined as the square root of the average of the squared differences between the predictions and the actual observations.

When interpreting PLSpredict results, the focus should be on the theoretical model's primary endogenous construct, and not the prediction errors for all endogenous constructs. When the primary endogenous construct has been selected, the Q^2_{predict} statistic should be evaluated first to verify that the predictions outperform the most naïve benchmark. The naïve value is produced when the indicator means from the analysis sample are used to predict the holdout sample (Shmueli et al., 2019). If the prediction results are better than the naïve value (above 0), researchers can then examine the other prediction statistics.

In most instances, researchers should use RMSE as the prediction statistic. But the MAE should be applied if the prediction error distribution is highly non-symmetrical (Shmueli et al., 2019). To assess the prediction error of a PLS-SEM analysis, the RMSE values are compared to a naïve value obtained by a linear regression model (LM) that generates predictions for the measured variables (indicators). The LM process applies a linear regression model that predicts each of the endogenous construct's indicators from all indicators of the exogenous latent variables in the PLS path model. But the LM process does not include the specified model structure represented by the measurement and structural theory (Danks & Ray, 2018).

The RMSE and MAE values are both acceptable prediction benchmarks, depending on the symmetry of the prediction error distribution. When interpreting the RMSE and MAE statistics, the values should be compared to the LM benchmark values. The following guidelines should be applied (Hair, Risher, et al., 2019; Shmueli et al., 2019):

- When the RMSE or MAE have higher prediction errors for all dependent variable indicators compared to the naïve LM benchmark, the model lacks predictive power.
- When the majority of the dependent construct indicators have higher prediction errors compared to the naïve LM benchmark, the model has low predictive power.

- When an equal or a minority of the dependent construct indicators have greater prediction errors compared to the naïve LM benchmark, the model has medium predictive power.
- When none of the dependent construct indicators have higher RMSE or MAE prediction errors compared to the naïve LM benchmark, the model has high predictive power.

Step 7: There are many options for advanced analysis when applying PLS-SEM. The advanced options include, for example, mediation, moderation, both categorical and continuous variables, multigroup analysis, invariance, unobserved heterogeneity, nonlinear effects, and endogeneity. These advanced analyses are beyond the scope of this article, and we refer you to Hair, Hult, et al. (2017) and Hair et al. (2018) for an extended discussion of these topics. We recommend, however, that researchers strongly consider performing these types of robustness checks on their analyses. Robustness checks evaluate the manner in which PLS-SEM results differ when analysis decisions are altered, similar to sensitivity checks in meta-analyses (Hair et al., 2018; Sarstedt, Ringle, et al., 2019), and they can be used to support the validity of statistical conclusions. Examples of robustness checks include adding or removing variables, moderation, modeling nonlinear relationships, and assessing endogeneity as well as unobserved heterogeneity. Sarstedt, Ringle, et al. (2019) recently reviewed methods to perform three of these robustness checks that can aid future researchers.

5. Conclusions and implications

To achieve the objectives of measurement model confirmation in developing or adapting multi-item measures, researchers could use either CCA or CFA. The results are different, however, and researchers need to understand the implications of the distinct outcomes so they can make informed decisions. CCA and CFA can both be used to improve individual item and scale reliability, identify and provide an indication of items that need to be revised or in some instances eliminated for content validity, facilitate achieving convergent validity and discriminant validity, and remove error variance. Compared to CFA, CCA has several benefits, including the following: (1) the number of items retained to measure constructs is higher with CCA, thereby improving construct validity, (2) determinant construct scores are available (Rigdon et al., 2019), (3) CCA can be applied to formative measurement models, and (4) when prediction is the statistical objective of the research, CCA as a component of PLS-SEM is the preferred method (Hair, Matthews, et al., 2017). Given these notable benefits, certain considerations of CCA should be emphasized.

CCA provides valuable implications beyond EFA, for both principal components analysis and common factor analysis. First, EFA is exploratory whereas CCA is confirmatory. While exploratory research is needed in certain circumstances, such as early in the scale development process, researchers utilizing the PLS-SEM approach can perform confirmatory analyses of reflective and formative composite structures via CCA, which broadens the applicability of both PLS-SEM and CCA. Second, CCA can move researchers beyond investigations of factor structures, and the analysis can assess the reliability, convergent validity, discriminant validity, and predictive validity of the measures. CCA thereby provides a more holistic perspective of measurement properties than EFA, and it can identify a wider range of measurement concerns when assessing the applicability of measures. Third, CCA can facilitate the assessment of reflective as well as formative measurement models, which is not easily achieved, and likely not possible, with traditional EFA approaches. A growing body of research is beginning to conceptualize and operationalize formative measures (Diamantopoulos, Riefler, & Roth, 2008; Fassott & Henseler, 2015), and CCA can play a pivotal role in this continued research growth. Lastly, variance extracted from exogenous constructs in CCA is specifically focused on the prediction of endogenous constructs, which poses important implications for model development and revision, as detailed above.

CCA also provides further alternatives regarding prediction. OLS regression is perhaps the most common method for examining causal relationships between exogenous and endogenous variables in the social sciences, but CCA is a superior approach when assessing prediction. The multi-item composites are weighted when CCA is applied as part of PLS-SEM, and the approach is therefore an improvement over using unweighted sum scores for independent variables in a multiple regression model. Indicators that are more strongly representative of composites are weighted more strongly when creating composite scores and assessing composite relationships, which provides a more accurate understanding of the underlying constructs of interest than traditional regression and EFA techniques. Finally, researchers should always use SEM methods, particularly PLS-SEM with CCA, when the measurement models are indirectly measured conceptual concepts and when examining mediation effects (Hair & Sarstedt, 2019; Nitzl, Roldán, & Cepeda, 2016). Therefore, CCA should always be considered as a technique when the focus of research is prediction.

We want to clarify that CCA and CFA are both acceptable approaches to develop and assess multi-item constructs. But researchers need to understand the differences between the two approaches and apply each technique when they are most appropriate. If researchers are focused on the content validity of constructs, whether reflective or formative measurement models are assumed, then CCA is a superior approach. CCA produces larger indicator loadings because the basis of developing solutions for PLS-SEM and CCA is total variance rather than common variance alone. These larger loadings result in the retention of a larger number of indicators, which produces more valid constructs. In conducting these analyses, researchers should also be aware that the guidelines and requirements regarding CFA do not necessarily apply to CCA – and vice versa. Notably, the execution of CCA within a PLS path model does not require the assessment of fit. CCA and PLS-SEM in general should be assessed based on the metrics unique to variance-based SEM, and goodness of fit is not a required metric. As discussed above, we recommend solutions developed using CCA and PLS-SEM should be evaluated based on predictive validity, particularly out-of-sample prediction obtained from PLSpredict.

The choice of CCA or CFA is deeply intertwined with the choice of PLS-SEM or CB-SEM. Guidelines are provided, therefore, relative to the application of these two structural modeling approaches. First, some researchers may simply prefer the study of composites, whereas others may prefer the study of factors. The ramifications of studying composites or factors are summarized above, and researchers can determine which approach is most appropriate. Second, PLS-SEM is often associated with exploration and development of theory, whereas CB-SEM is most often associated with confirming theory (Hair, Black, et al., 2019; Hair et al., 2018). PLS-SEM is more useful in the earlier phases of theory development, while CB-SEM may be more useful in the latter phases, assuming prediction is not the objective of the research. Third, PLS-SEM provides more accurate estimates with small sample sizes, and it should therefore be applied in such instances (Hair & Sarstedt, 2019). Fourth, PLS-SEM is more likely to result in model convergence when studying a large number of observed and/or latent variables, and it is more appropriate when models are complex (Hair, Black, et al., 2019; Hair et al., 2018). Fifth, PLS-SEM should be chosen when prediction is a primary focus of the research (Shmueli et al., 2016, 2019). CB-SEM is not the most effective structural modeling method for predictive models. Sixth, CB-SEM provides traditional indicators of model fit, and should be applied when such indicators are needed. Seventh, researchers should be aware a variety of statistical questions were previously only possible to be addressed via CB-SEM. But recent advancements enable PLS-SEM to likewise address these questions. For instance, endogeneity and measurement invariance are easily assessed with PLS-SEM. Finally, other options not possible or at minimum difficult with CB-SEM are easily executed with PLS-SEM, such as continuous moderators, mediation and multi-group analysis. Thus, researchers should regularly review new publications regarding PLS-SEM to apply the most appropriate method for their research questions.

This article summarizes important new concepts for researchers to consider when developing and evaluating the quality of measurement models. The concepts are also important for reviewers and journal editors to ensure the latest approaches are applied in published PLS-SEM studies. We summarize an emerging role for CCA, an alternative to the application of CFA in the development, adaptation and confirmation of measurement scales. Finally, while a few articles have been published that are negative about the use of PLS-SEM, more recently several prominent researchers have acknowledged the value of PLS as an SEM technique (Petter, 2018). We believe that social science scholars would be remiss if they did not apply all statistical methods at their disposal to explore and better understand the phenomena they are researching.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbusres.2019.11.069.

References

Becker, J.-M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type models. *Long Range Planning, 45*(5–6), 359–394.

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2019). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*. https://doi.org/10.1016/j.im.2019.05.003.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25*(1), 67–74.

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research.* New York, NY: Guilford Publications.

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly, 33*(4), 689–708.

Cheah, J.-H., Sarstedt, M., Ringle, C. M., Ramayah, T., & Ting, H. (2018). Convergent validity assessment of formatively measured constructs in PLS-SEM: On using single-item versus multi-item measures in redundancy analyses. *International Journal of Contemporary Hospitality Management, 30*(11), 3192–3210.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.). *Modern Methods for Business Research* (pp. 295–358). Mahwah: Erlbaum.

Chin, W. W. (2003). *PLS graph 3.0.* Houston: Soft Modeling Inc.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Cohen, J. (1994). The world is round (p < . 05). *American Psychologist, 49*(12), 997–1003.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando, Florida: Holt, Rinehart and Winston.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302.

Danks, N. P., & Ray, S. (2018). Predictions from partial least squares models. In F. Ali, S. Rasoolimanesh, & C. Cobanoglu (Eds.). *Applying partial least squares in tourism and hospitality research* (pp. 35–52). Bingley: Emerald Publishing Limited.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203–1218.

Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis, 81*(1), 10–23.

Evermann, J., & Tate, M. (2016). Assessing the predictive performance of structural equation model estimators. *Journal of Business Research, 69*(10), 4565–4582.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods, 4*(3), 272–299.

Fassott, G., & Henseler, J. (2015). *Formative (measurement) Wiley Encyclopedia of Management, Vol. 9.* Wiley.

Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: A comparison of four procedures. *Internet Research, 29*(3), 430–447.

Geisser, S. (1974). A predictive approach to the random effects model. *Biometrika, 61*(1), 101–107.

Goetz, C., Coste, J., Lemetayer, F., Rat, A.-C., Montel, S., Recchia, S., ... Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical Epidemiology, 66*(7), 710–718.

Grimm, L. G., & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Guilford, J. P. (1936). *Psychometric methods.* New York: McGraw-Hill.

Hair, J. F., Black, W. C., Anderson, R. E., & Babin, B. J. (2019). *Multivariate data analysis* (8th ed.). London: Cengage Learning.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Thousand Oaks: Sage.

Hair, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: Updated guidelines on which method to use. *International Journal of Multivariate Data Analysis, 1*(2), 107–123.

Hair, J. F., Page, M., & Brunsveld, N. (2020). *Essentials of business research methods* (4th ed.). New York, NY: Routledge.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–151.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2012). Partial least squares: The better approach to structural equation modeling? *Long Range Planning, 45*(5–6), 312–319.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24.

Hair, J. F., & Sarstedt, M. (2019). Factors vs. Composites: guidelines for choosing the right structural equation modeling method. *Project Management Journal* Forthcoming.

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2019). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing, 53*(4), 566–584.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced issues in partial least squares structural equation modeling.* Thousand Oaks: SAGE Publications.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science, 40*(3), 414–433.

Hakel, M. D. (2000). Dr. G. Frederic Kuder 1903–2000. *Personnel Psychology, 53*(2), 273–274.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Press.

Henseler, J., & Dijkstra, T. K. (2015). ADANKO 2.0. Composite Modeling. Kleve, Germany.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., ... Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö & Evermann (2013). *Organizational Research Methods, 17*(2), 182–209.

Henseler, J., Hubona, G. S., & Pauline, A. R. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems, 116*(1), 2–20.

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.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33* forthcoming.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction, 32*(1), 51–62.

Howard, M. C. (2018). Scale pretesting. *Practical Assessment, Research & Evaluation, 23*(5).

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika, 69*(1), 81–99.

Jöreskog, K. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: historical and comparative aspects. In H. Wold, & K. G. Jöreskog (Eds.). *Systems under indirect observation, Part I* (pp. 263–270). Amsterdam: North-Holland.

Kline, P. (2014). *An easy guide to factor analysis.* London: Routledge.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Publications.

Kuppelwieser, V. G., & Sarstedt, M. (2014). Applying the future time perspective scale to advertising research. *International Journal of Advertising, 33*(1), 113–136.

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares.* Heidelberg: Physica.

Lorr, M., & Heiser, R. B. (1965). Marion webster Richardson (1896–1965). *Psychometrika, 30*(3), 235–237.

Mateos-Aparicio, G. (2011). Partial least squares (PLS) methods: Origins, evolution, and application to social sciences. *Communications in Statistics - Theory and Methods, 40*(13), 2305–2317.

Nitzl, C. (2016). The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development. *Journal of Accounting Literature, 39*, 19–35.

Nitzl, C., & Chin, W. W. (2017). The case of partial least squares (PLS) path modeling in managerial accounting research. *Journal of Management Control, 28*(2–3), 137–156.

Nitzl, C., Roldán, J. L., & Cepeda, G. (2016). Mediation analyses in partial least squares structural equation modeling: Helping researchers to discuss more sophisticated models. *Industrial Management & Data Systems, 116*(9), 1849–1864.

Petter, S. (2018). "Haters Gonna Hate": PLS and information systems research. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 49*(2), 10–13.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 2*(1), 13–43.

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis.* New Jersey, NJ: Wiley.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling. In praise of simple methods. *Long Range Planning, 45*(5–6), 341–358.

Rigdon, E. E. (2014). Rethinking partial least squares path modeling: Breaking chains and forging ahead. *Long Range Planning, 47*, 161–167.

Rigdon, E. E., Becker, J.-M., & Sarstedt, M. (2019). Factor indeterminacy as metrological uncertainty: Implications for advancing psychological measurement. *Multivariate Behavioral Research* forthcoming.

Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A critical look at the use of PLS-SEM in MIS quarterly. *MIS Quarterly, 36*(1), iii–xiv.

Ringle, C. M., Wende, S., & Becker, J.-M. (2014). SmartPLS 3: www.smartpls.de.
Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS 2.0. Hamburg: www.smartpls.de.
Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies!. *Journal of Business Research, 69*(10), 3998–4010.
Sarstedt, M., Hair, J. F., Jr, Cheah, J.-H., Becker, J.-M., & Ringle, C. M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal* in press.
Sarstedt, M., Ringle, C. M., Cheah, J.-H., Ting, H., Moisescu, O. I., & Radomir, L. (2019). Structural model robustness checks in PLS-SEM. *Tourism Economics* 1354816618823921.
Sarstedt, M., Ringle, C. M., Henseler, J., & Hair, J. F. (2014). On the Emancipation of PLS-SEM: A commentary on Rigdon (2012). *Long Range Planning, 47*(3), 154–160.
Schuberth, F., Henseler, J., & Dijkstra, T. K. (2018). Confirmatory composite analysis. *Frontiers in Psychology, 9*, 2541.
Sharma, P., Sarstedt, M., Shmueli, G., Kim, K. H., & Thiele, K. O. (2019). PLS-based model selection: The role of alternative explanations in information systems research. *Journal of the Association for Information Systems, 20*(4).
Sharma, P., Shmueli, G., Sarstedt, M., Danks, N., & Ray, S. (2019). Prediction-oriented model selection in partial least squares path modeling. *Decision Sciences, in press.*. https://doi.org/10.1111/deci.12329.
Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552–4564.
Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*. https://doi.org/10.1108/EJM-02-2019-0189.
Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology, 15*(2), 201–293.
Stevens, J. (2009). *Applied multivariate statistics for the social sciences.* Taylor & Francis US.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, 36*(2), 111–147.
Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen, J. A. (1950). *Measurement and prediction, Vol. 4*. Princeton, NJ: Princeton University Press.
Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications.* Washington, DC: American Psychological Association.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70.
Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog, & H. Wold (Eds.). *Systems under indirect observations: Part II* (pp. 1–54). Amsterdam: North-Holland.
Wood, M. (2005). Bootstrapped confidence intervals as an approach to statistical inference. *Organizational Research Methods, 8*(4), 454–470.
Yule, G. U. (1907). On the theory of correlation for any number of variables, treated by a new system of notation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 79*(529), 182–193.
Yule, G. U. (1911). *An introduction to the theory of statistics.* London: Griffin.

Joe Hair is Director of the PhD Program and Cleverdon Chair of Business, Mitchell College of Business, the University of South Alabama, Mobile, Alabama. Clarivate Analytics recently recognized him for being in the top 1% globally of all Business and Economics professors based on his citations and scholarly accomplishments, which for his career exceed 196,000. He has authored over 80 book editions, including *Multivariate Data Analysis*, Cengage Learning, U.K., 8th edition, 2019 (cited 118,000+ times and one of the top five all time social sciences research methods textbooks); *MKTG*, Cengage Learning, 13th edition, 2020; *Essentials of Business Research Methods*, Routledge, 4th edition 2020; *Essentials of Marketing Research*, McGraw-Hill, 5th edition 2020; and *A Primer on Partial Least Squares Structural Equation Modeling*, Sage, 2nd edition 2017. He also has published numerous articles in scholarly journals such as the *Journal of Marketing Research*, *Journal of Academy of Marketing Science*, *Organizational Research Methods*, *Journal of Advertising Research*, *Journal of Business Research*, *Journal of Long Range Planning*, *Industrial Marketing Management*, *European Journal of Marketing, European Journal of Management, Journal of Retailing*, and others. He is writing a new book on *Marketing Analytics*, forthcoming in 2020 (McGraw-Hill).

Matt C. Howard is Assistant Professor, Mitchell College of Business, at the University of South Alabama, Mobile, Alabama. His research typically focuses on (a) statistics and research methodologies, (b) employee-computer interactions, (c) training and development, and (d) personality. His research has appeared in the *Journal of Applied Psychology, Organizational Research Methods, Journal of Organizational Behavior, Information & Management, Human-Computer Interaction*, and the *Journal of Business Research*.

Christian Nitzl is on the faculty of the Center for Intelligence and Security Studies, University of the German Federal Armed Forces, Munich. Germany. His research interests include PLS path modelling, with a special focus on the application of PLS-SEM in accounting, trust research, and accounting change in the governmental area. He has published in Financial Accountability & Management, Journal of Accounting Literature, Journal of Accounting & Organizational Change, Journal of Banking & Finance, Journal of Public and Nonprofit Management, Public Management Review, and Die Betriebswirtschaft, among others. He has served many times as a reviewer for journals. Furthermore, he regularly coaches researchers and practitioners about the application and interpretation of PLS-SEM.