



ELSEVIER

Contents lists available at ScienceDirect

Journal of Business Research

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

Using confirmatory composite analysis to assess emergent variables in business research ☆, ☆ ☆

Jörg Henseler^{a,b,*}, Florian Schubert^a^a University of Twente, Faculty of Engineering Technology, Drienerlolaan 5, 7522 NB Enschede, the Netherlands^b Universidade Nova de Lisboa, Nova Information Management School, Campus de Campolide, Lisbon, Portugal

ARTICLE INFO

Keywords:

Confirmatory composite analysis
CCA
Composite model
Emergent variables
Structural equation modeling
Covariance structure analysis

ABSTRACT

Confirmatory composite analysis (CCA) was invented by Jörg Henseler and Theo K. Dijkstra in 2014 and elaborated by Schubert et al. (2018b) as an innovative set of procedures for specifying and assessing composite models. Composite models consist of two or more interrelated constructs, all of which emerge as linear combinations of extant variables, hence the term ‘emergent variables’. In a recent JBR paper, Hair et al. (2020) mistook CCA for the measurement model evaluation step of partial least squares structural equation modeling. In order to clear up potential confusion among JBR readers, the paper at hand explains CCA as it was originally developed, including its key steps: model specification, identification, estimation, and assessment. Moreover, it illustrates the use of CCA by means of an empirical study on business value of information technology. A final discussion aims to help analysts in business research to decide which type of covariance structure analysis to use.

1. Introduction

Modeling and testing theories that contain abstract concepts is a core part of a large number of studies in business research. The family of statistical tools provided for this purpose, covariance structure analysis (CSA), plays a crucial role in this process. Methodological papers on CSA such as Bagozzi and Yi (1988) or Fornell and Larcker (1981) are among the most cited papers in marketing and related fields. An important task in CSA is the operationalization of these abstract concepts (Sajtos & Magyar, 2016). For a long time, constructs—the representations of concepts in the statistical model—have been equated with latent variables, which typically are common factors that fully explain the covariation among the observable variables (Borsboom, 2008). The preferred type of CSA to assess this type of latent variable model is confirmatory factor analysis (CFA).

Due largely to the poor test record of the common factor model (Henseler et al., 2014), researchers are looking for alternatives to latent variables. Rigdon (2012, p. 342) observed that “research in statistics and psychometrics challenges the factor-centric worldview,” and Rhemtulla, van Bork, and Borsboom (2020) recommended the exploration of opportunities offered by composite methods. An alternative

type of construct is found in emergent variables (Cole, Maxwell, Arvey, & Salas, 1993). An emergent variable is a composite of variables of which the correlations with other variables in a model are proportional to one another (Benitez, Henseler, Castillo, & Schubert, 2020; Dijkstra, 2017). Alternative terms for emergent variables are ‘composite constructs’ (Benitez, Llorens, & Braojos, 2018), ‘aggregate constructs’ (Edwards, 2001), and ‘formative constructs’² (Petter, Straub, & Rai, 2007). Emergent variables have been used much less frequently than latent variables; possibly because of the lack of statistical tools to assess them. Empirical assessment is as important for emergent variables as it is for latent variables. As Rigdon (2012, p. 353) explained, “[r]ejecting the factor model does not mean rejecting rigor, but it does mean defining rigor in composite terms.” Analysts who would like to employ emergent variables in their models therefore need a statistical method to assess them. This is where confirmatory composite analysis (CCA) enters the stage.

CCA, invented by Jörg Henseler and Theo K. Dijkstra (see the author note in Henseler et al., 2014), is an innovative set of procedures for specifying and assessing composite models. It is a fully developed method for confirmatory purposes that assesses composite models with the same rigor as CFA does for common factor models (Schubert,

* This paper reuses material from Schubert et al.’s original publication on confirmatory composite analysis: <https://doi.org/10.3389/fpsyg.2018.02541>.

** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author at: University of Twente, Faculty of Engineering Technology, Drienerlolaan 5, 7522 NB Enschede, the Netherlands.

E-mail addresses: j.henseler@utwente.nl (J. Henseler), f.schubert@utwente.nl (F. Schubert).

¹ The first author acknowledges a financial interest in ADANCO and its distributor, Composite Modeling.

² Note that there is also formative measurement of latent variables, which we do not mean here.

<https://doi.org/10.1016/j.jbusres.2020.07.026>

Received 5 February 2020; Received in revised form 12 July 2020; Accepted 18 July 2020

0148-2963/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Henseler, & Dijkstra, 2018b). In fact, CCA has been designed analogous to CFA. The only difference between the two is that whereas CFA helps to assess a latent variable structure of observable variables, CCA helps to assess an emergent variable structure. Similar to all types of CSA, CCA examines the discrepancy between the empirical and the model-implied variance–covariance matrix of observable variables, i.e., the model's goodness of fit. As Barrett (2007, p. 823) has emphasized, “model fit testing and assessment is paramount, indeed crucial, and cannot be fudged for the sake of ‘convenience’ or simple intellectual laziness on the part of the investigator.”

In a recent JBR paper, Hair, Howard, and Nitzl (2020) claimed to “introduce and explain the process of CCA” (p. 106). However, not only did they forget to mention at the beginning from whom they had borrowed the idea of CCA; they also confused CCA with the measurement model evaluation step in ‘partial least squares structural equation modeling’ (PLS-SEM), an approach promoted as a “silver bullet” by Hair, Ringle, and Sarstedt (2011), which is “very appealing to many researchers” (Hair, Risher, Sarstedt, & Ringle, 2019, p. 3).³ Due to this confusion, they made several errors regarding CCA as Schuberth et al. (2018b) proposed it. For instance, according to Hair et al. (2020, p. 108): (1) “CCA should always be considered as a technique when the focus of research is prediction,” (2) “CCA can facilitate the assessment of reflective as well as formative measurement models,” and (3) “goodness of fit is not a required metric.” In all three aspects, Hair et al. (2020) obviously erred. As its name indicates, CCA as proposed by Schuberth et al. (2018b) is dedicated to confirmation (actually disconfirmation), i.e., to assessing whether a proposed composite model fits the data, and not to predicting outcomes. Also, its focus is on composite models comprising emergent variables and not on reflective or formative measurement models consisting of latent variables. Finally, as a form of CSA, CCA centrally contains the concept of model fit assessment. Evidence of model fit is a necessary but not sufficient condition for a model to be true. For a more elaborate comparison of CCA proposed by Schuberth et al. (2018b) and Hair et al. (2020), we refer to Schuberth (in press).

Against the background given above, the study at hand expounds the confusion regarding CCA introduced by Hair et al. (2020), presents CCA as originally developed, explains its steps, illustrates its use, and discusses when it should be applied. The next section presents CCA as it was first introduced by Henseler et al. (2014) and Schuberth et al. (2018b), i.e., as a type of CSA to specify, estimate, and assess composite models. The following section illustrates how a CCA is performed using a case from the domain of information technology's business value and discusses potential limitations. The final section provides conclusions as well as guidelines for analysts to make well-motivated decisions on which models and methods to employ.

2. Confirmatory composite analysis

For a long time, CSA, particularly CFA and structural equation modeling (SEM), was mainly equated with latent variable modeling, i.e., abstract concepts are operationalized by means of latent variables. Rigdon (2012) wrote a groundbreaking article in which he questioned the factual monopoly of the common factor model and proposed the so-called *concept proxy framework*, which provides a fresh perspective on the gap between a concept and its corresponding observable variables. This framework opened the way for employing emergent variables to approximate abstract concepts. Emergent variables are defined by their indicators (Reise, 1999). As linear combinations of other variables, they fulfill the requirement of composite models to act along a single

³ Taking into account that Hair was a co-author in Henseler et al. (2014) as well as a listener to Henseler (2015a), it is even more surprising that Hair et al. (2020) dubbed the measurement model evaluation step in PLS-SEM CCA and created unnecessary confusion.

dimension (Dijkstra, 2017). They can be described as “conceptual conveniences—[...] phenomena that exist at a higher level than their constituent elements for reasons that people find useful” (Coan, 2010, p. 278). Inspired by Rigdon's (2012) idea and following his postulate for composite-based SEM to show the same rigor as factor-based SEM, the developers of CCA designed their method as a new type of CSA that functions to assess composite models instead of common factor models.

CCA is analogous to CFA, but rather than a common factor model (also known as reflective measurement model), it comprises primarily a composite model. Hence, the abstract concept under investigation is represented by an emergent variable instead of a latent variable in the statistical model (Benitez et al., 2020; Chin, 2010; Cohen, Cohen, Teresi, Marchi, & Velez, 1990; Cole et al., 1993; Reise, 1999). Against this background, CCA is used to assess whether the constraints imposed by the composite model are consistent with the data, i.e., whether the information between two blocks of observable variables is fully conveyed by the emergent variables. Thus, by employing CCA, researchers can empirically assess their postulated theories expressed by statistical models containing emergent variables.

It took more than four years to develop CCA. While the initial idea of CCA was already sketched by Henseler et al. (2014), the scholarly publishing process took more time, until finally, Schuberth et al. (2018b) provided the first full description of the method and demonstrated CCA's efficacy by means of a Monte Carlo simulation. Interim developments on CCA were shared with the scientific community either in written form (Henseler, 2015c, 2017; Henseler, Hubona, & Ray, 2016) or in oral communication (Henseler, 2015a, 2015b, 2015d; Schuberth, Dijkstra, & Henseler, 2018a). A video presentation by Henseler (2015a) is available at <http://tv.us.es/videoembed/?numberpost=30277>.

Despite the fact that CCA is a very new method, it has already found widespread application across business disciplines, for instance in family business (Pittino, Martínez, Chirico, & Galván, 2018; Ruiz-Palomo, Diéguez-Soto, Duréndez, & Santos, 2019), information systems research (Benitez, Llorens, et al., 2018; Braojos, Benitez, & Llorens, 2019), innovation management (Cegarra-Navarro, Papa, Garcia-Perez, & Fiano, 2019; Cegarra-Navarro, Ruiz, Martínez-Caro, & Garcia-Perez, in press), knowledge management (Martelo-Landroguez, Cegarra Navarro, & Cepeda-Carrión, 2019; Roldán, Real, & Sánchez Ceballos, 2018; Sánchez-Polo, Cegarra-Navarro, Cillo, & Wensley, 2019), management education (Hernández-Perlines, Moreno-García, & Yáñez-Araque, 2016; Rueda, Benitez, & Braojos, 2017), marketing (Foltean, Trif, & Tuleu, 2019), organizational behavior (Felipe, Roldán, & Leal-Rodríguez, 2016), project management (Benítez-Ávila, Hartmann, Dewulf, & Henseler, 2018), service management (Yiu, Ngai, & Lei, 2020), supply chain management (Wei, Ke, Liu, & Wei, 2020), and tourism management (Rasoolimanesh, Md Noor, Schuberth, & Jaafar, 2019; Sanchez-Franco, Cepeda-Carrion, & Roldán, 2019).

As Fig. 1 illustrates, CCA consists of the same four steps as CFA and other forms of CSA: (1) model specification, (2) model identification, (3) model estimation, and (4) model assessment. Each of these steps is presented below in separate subsections.

2.1. Specifying composite models

In CCA, each emergent variable η_j is assumed to be composed of a unique block of K_j observable variables, $\mathbf{y}_j' = (y_{j1} \dots y_{jK_j})$, $\eta_j = \mathbf{w}_j' \mathbf{y}_j$.⁴ Thus, the composite model satisfies the principle that all information between the blocks of observable variables is conveyed solely by the emergent variables (Dijkstra, 2017). This principle entails that the

⁴ In general, also emergent variables composed of latent variables and emergent variables composed of emergent variables are conceivable, see for instance van Riel, Henseler, Kemény, and Sasovova (2017) and Schuberth, Rademaker, and Henseler (in press), respectively.

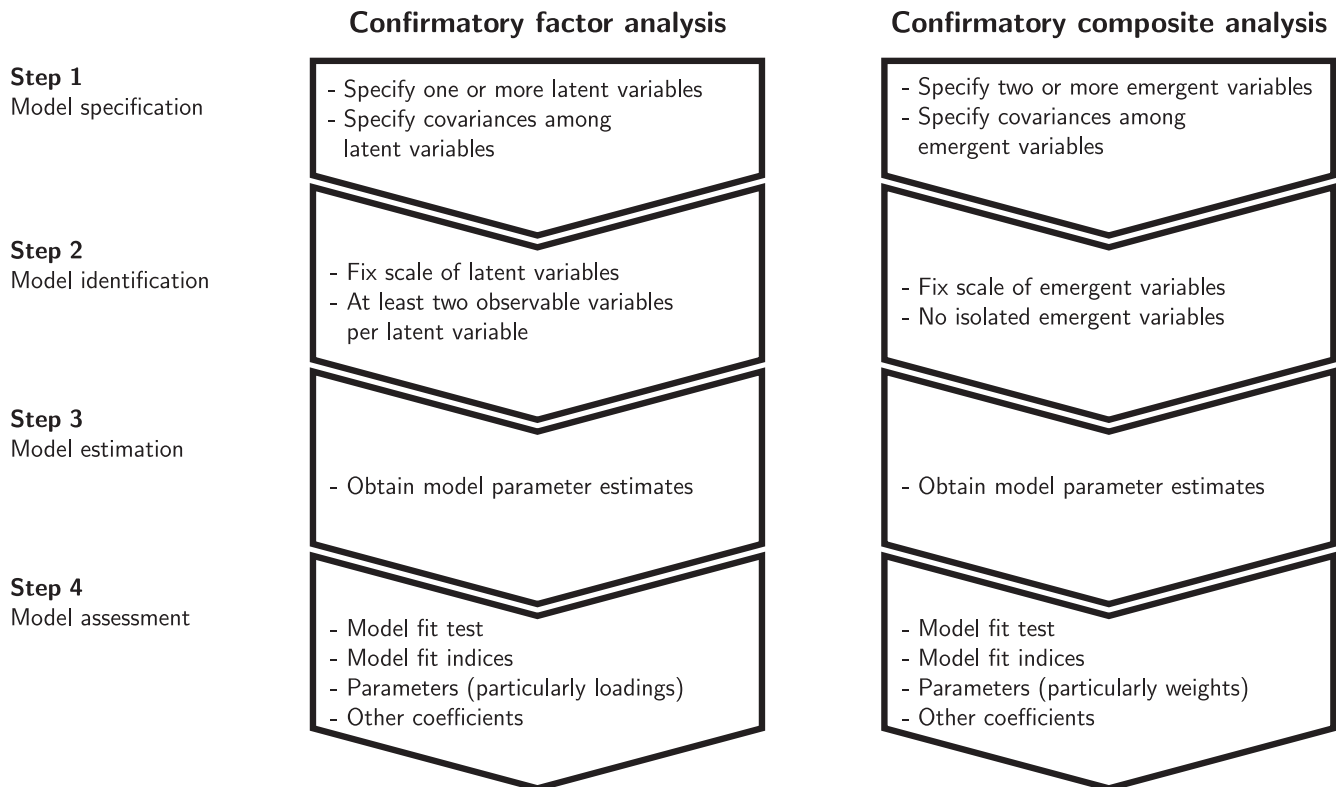


Fig. 1. A comparison of the Steps of CFA and CCA.

covariance between two observable variables belonging to different blocks is constrained to be the product of three factors: the two covariances between the observable variables and their respective emergent variable, and the covariance between the two emergent variables. In general, constraints on the variance–covariance matrix of the emergent variables are possible, and the emergent variables can be embedded in a structural model similar to SEM with latent variables (Dijkstra, 2017). That means Dijkstra (2017) envisioned SEM with emergent variables. For simplicity going further, we assume that all emergent variables freely covary, i.e., all covariances among the emergent variables are model parameters that are freely estimated.

The weights of block j to form the emergent variable η_j are captured in the vector \mathbf{w}_j of length K_j . Usually, each weight vector is scaled to ensure that the emergent variables have unit variance (see also the conditions for model identification as formulated in Section 2.2). Moreover, it is assumed that each observable variable is connected to only one emergent variable. The model-implied variance–covariance matrix Σ of the observable variables can be expressed as a partitioned matrix, as Eq. 1 shows:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1J} \\ & \Sigma_{22} & \cdots & \Sigma_{2J} \\ & & \ddots & \vdots \\ & & & \Sigma_{JJ} \end{pmatrix} \quad (1)$$

The intra-block variance–covariance matrix Σ_{jj} of dimension $K_j \times K_j$ is typically unconstrained, and captures the covariation among the observable variables of block j ; thus, the observable variables of one block can freely covary. Moreover, it can be shown that the variance–covariance matrix of the observable variables is positive-definite if and only if the following two conditions hold (Dijkstra, 2015, 2017):

- (i) all intra-block variance–covariance matrices are positive-definite, and

- (ii) the variance–covariance matrix of the emergent variables is positive-definite.

The covariances between the observable variables of block j and l are captured in the inter-block covariance matrix Σ_{jl} , with $j \neq l$ of dimension $K_j \times K_l$. However, in contrast to the intra-block variance–covariance matrix, the inter-block covariance matrix is constrained, since by assumption, the emergent variables carry all information between the blocks:

$$\Sigma_{jl} = \rho_{jl} \Sigma_{jj} \mathbf{w}_j \mathbf{w}_l' \Sigma_{ll} = \rho_{jl} \lambda_j \lambda_l', \quad (2)$$

where ρ_{jl} presents the covariance between the emergent variables η_j and η_l . The vector $\lambda_j = \Sigma_{jj} \mathbf{w}_j$ of length K_j contains the loadings, which are defined as the covariances between the emergent variable η_j and its associated observable variables \mathbf{y}_j . Eq. 2 is highly reminiscent of the corresponding equation in CFA, in which all concepts are modeled as latent variables instead of emergent variables. In the classical reflective measurement model, the vector λ_j captures the factor loadings of the observable variables on their connected latent variable, and ρ_{jl} represents the covariance between latent variables j and l . Hence, both models show the rank-one structure for the covariance matrices between two blocks of observable variables.

Although the intra-block variance–covariance matrices of the observable variables Σ_{jj} are typically not constrained, importantly, the composite model is still a model from the CSA point of view. It assumes that all information between the observable variables of two different blocks is conveyed by the emergent variable(s), and therefore, it imposes rank-one restrictions on the inter-block covariance matrices of the observable variables (see Eq. 2). These restrictions can be exploited to assess the overall model fit (see Section 2.4). Notably, the weights \mathbf{w}_j producing these matrices are the same across all inter-block covariance matrices Σ_{jl} with $l = 1, \dots, J$ and $l \neq j$.

To specify a composite model in CCA, the researcher has to decide

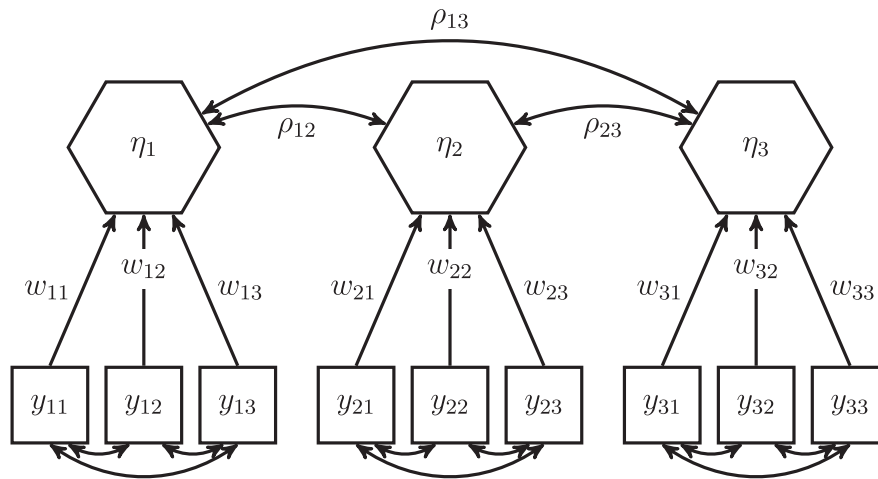


Fig. 2. A composite model consisting of three correlated emergent variables.

on the covariances among the emergent variables and the observable variable forming these emergent variables. Fig. 2 shows an exemplary composite model consisting of three interrelated emergent variables, each of which is composed of three observable variables.

The composite model typically allows the observable variables of each emergent variable to be freely correlated, which is indicated by the double-headed arrows among the observable variables of one block. Similarly, the three emergent variables are allowed to freely correlate, which is highlighted by the double-headed arrows among the three emergent variables. Finally, the correlations between the observable variables of two different emergent variables are fully conveyed by the emergent variables.

2.2. Identifying composite models

As in SEM and CFA, model identification plays a preponderant role in CCA. Since analysts can freely specify their models, it must be ensured that the model parameters have a unique solution (Bollen, 1989, Chap. 8). Model identification is also necessary to obtain consistent parameter estimates and to reliably interpret them (Marcoulides & Chin, 2013; Martín & Quintana, 2002). For composite models, as those studied in CCA, for identification, at least two conditions must be fulfilled (Dijkstra, 2017; Schubert et al., 2018b). Each condition is necessary but not sufficient for identification.

First, a necessary condition for ensuring identification is to fix the scale of each emergent variable. This can be done by either fixing one weight per emergent variable or fixing the variance of each emergent variable. Typically, the weights are chosen to ensure that the variance of each emergent variable is one, $w_j' \Sigma_j w_j = 1$. If this approach is applied, the sign of each weight vector of every block also needs to be determined, because the negative weight vector also leads to unit variance of the emergent variable. This is similar to a reflective measurement model if the variance of the latent variable is fixed as one.

Second, each emergent variable must be connected to at least one other variable not part of this emergent variable. In principle, the other variable can be observable, latent, or emergent. As a result, at least one inter-block covariance matrix Σ_{jl} , $l = 1, \dots, J$ with $l \neq j$ satisfies the rank-one condition. In normalizing the weight vectors, all model parameters can be uniquely retrieved from the variance–covariance matrix of observable variables since there is a non-zero inter-block covariance matrix for every weight vector. Otherwise, if an emergent variable η_j is isolated, all inter-block covariance matrices with $l \neq j$, belonging to this emergent variable are of rank zero. Consequently, the

weights forming this emergent variable cannot be uniquely retrieved, because an infinite number of weight sets exists for this emergent variable that satisfies the normalization condition.

2.3. Estimating composite models

The existing literature provides various methods to construct emergent variables from blocks of observable variables. The most common among them are principal component analysis (PCA, Pearson, 1901), linear discriminant analysis (LDA, Fisher, 1936), generalized canonical correlation analysis (GCCA, Kettenring, 1971), and the iterative partial least squares algorithm (PLS, Wold, 1973). All these approaches seek emergent variables that ‘best’ explain the data and can be regarded as prescriptions for dimension reduction (Dijkstra & Henseler, 2011).

In their original article on CCA, Schubert et al. (2018b) used MAXVAR, an approach to GCCA, to consistently estimate composite models. Although it is suggestive to apply a composite-based estimator such as the iterative PLS algorithm or generalized structured component analysis (GSCA, Hwang & Takane, 2004), there is no obvious reason not to use other estimators, such as maximum likelihood (ML) or generalized least squares. The decision to use a certain estimator should be based on the estimator’s statistical properties and the validity of the assumptions under which these properties were derived, such as independent and identically distributed (i.i.d.) random variables. Typically, unbiased and/or consistent estimators are favored over biased and/or inconsistent ones. Purely practical aspects such as computation time and convergence behavior tend to play a minor role in the selection of estimators. Finally, although not further discussed here, Bayesian estimation is also conceivable (see e.g., Choi & Hwang, 2020; Vidaurre, van Gerven, Bielza, Larrañaga, & Heskes, 2013).

In cases where researchers face a high degree of multicollinearity, it can be beneficial to avoid consistent estimators and choose an estimator that is robust against multicollinearity, but inconsistent. This also applies to the iterative PLS algorithm, where the weights obtained by PLS Mode B are consistent but affected by multicollinearity, while the weights obtained by PLS Mode A are usually inconsistent but unaffected by high correlations among the observable variables. In such a situation, employing Mode A could be advantageous even if the obtained weight estimates are generally not expected to be consistent. An analogous approach is customary in the context of regression analysis where ridge regression is employed in situations with a high multicollinearity among the independent variables (Mason & Brown, 1975).

2.4. Assessing composite models

Model assessment is a pivotal step in CCA. In principle, it does not differ from assessing structural equation models generally. In CCA, it consists of the following two steps: (i) overall model fit assessment, and (ii) inspection of each emergent variable separately.

As in SEM and CFA, overall model fit assessment is crucial in CCA (Mulaik et al., 1989; Schubert et al., 2018b; Yuan, 2005). At the heart of overall model fit assessment is the discrepancy between the empirical and the model-implied variance–covariance matrix of the observable variables. This discrepancy can be exploited to obtain empirical evidence against the specified model, i. e., we examine whether the constraints imposed by the model are justifiable. Note that “[i]f a model is consistent with reality, then the data should be consistent with the model. But if the data are consistent with the model, this does not imply that the model corresponds to reality” (Bollen, 1989, p. 68).

Schubert et al. (2018b) proposed two nonexclusive ways to assess composite models: (i) a bootstrap-based test for exact overall model fit (Beran & Srivastava, 1985) and (ii) fit indices. The goal of the bootstrap-based test for exact overall model fit is to assess the null hypothesis that the model-implied variance–covariance matrix based on the population parameters equals the population variance–covariance matrix of the observable variables: $H_0: \Sigma(\theta) = \Sigma$. The assessment relies on bootstrap to obtain the reference distribution of the discrepancy between the empirical variance–covariance matrix of the observable variables and their estimated model-implied counterpart under the null hypothesis of exact model fit. Hence, the bootstrap-based test is non-parametric in nature and functions asymptotically well under rather mild assumptions such as i.i.d. random variables (Beran & Srivastava, 1985). To measure this discrepancy, various metrics can be employed. In the context of CCA, the standardized root mean square residual (SRMR), the squared Euclidean distance, and the geodesic distance have been proposed (Schubert et al., 2018b).

To overcome the critiques of the test's stringency for exact overall model fit, fit indices such as the SRMR (Hu & Bentler, 1998) or the root mean square error of approximation (RMSEA, see Browne & Cudeck, 1993) have been proposed.⁵ Instead of defining exact fit as the desirable objective, fit indices quantify the discrepancy between the model and the data along a continuous scale to examine how well the estimated model fits the collected data. To judge whether a model shows an acceptable fit, the value of a fit index is typically compared to threshold values derived by simulation studies.

In empirical research, scientists are encouraged to examine alternative explanations for a phenomena under investigation (Nuzzo, 2015). In the context of SEM, this inevitably leads to situations where researchers face various alternative models that are theoretically plausible, i. e., guided by existing theory. To cope with this issue, considering model selection criteria can be helpful to choose the “optimal” model among alternative models. In this context, ‘optimal’ refers to the trade-off between model fit and model parsimony (Huang, 2017). The most prominent model selection criteria are arguably Akaike's information criterion (AIC Akaike, 1998) and the Bayesian information criterion (BIC Schwarz, 1978); however, various extensions such as the consistent AIC (Bozdogan, 1987) have been developed. For an overview of model selection criteria, we refer to McQuarrie and Tsai (1998) and West, Taylor, and Wu (2012). Notably, model selection criteria such as the BIC and the AIC, are based on model parameter estimates derived from the collected sample. Hence, similar to parameter estimates, model selection criteria are also subject to sampling variation (known as model selection uncertainty, see e.g., Burnham & Anderson, 2002),

⁵ For an overview of various fit indices, we refer to Schermelleh-Engel, Moosbrugger, and Müller (2003). However, it is noted that research investigating the performance of fit indices in the context of composite models is scarce.

which could result in different optimal models if a new sample from the same population is considered. To obtain more certainty on the optimality of a chosen model, replication of prior research is recommended (Preacher & Merkle, 2012).

Once a researcher has decided on a certain model, he/she typically continues to assess each emergent variable and its relation to its observable variables. In doing so, the weight estimates, loading estimates, and their significance are examined and compared to the expectations implied by a researcher's theory. Moreover, if weight estimates potentially suffer from multicollinearity, e.g., if PLS using Mode B is employed, multicollinearity needs to be assessed as well. Finally, the researcher needs to investigate whether the correlations among the emergent variables are in line with their expectations. For a more detailed description on how to assess emergent variables locally in the context of PLS path modeling, we refer to Benitez et al. (2020) and Henseler et al. (2016).

3. A demonstration of CCA

To illustrate the application of CCA, we refer to aspects of a CCA that Benitez, Ray, and Henseler (2018) conducted and reported.⁶ Their study in the domain of information technology's (IT's) business value investigated whether IT infrastructure flexibility has an impact on the success or failure of mergers and acquisitions. It followed a two-step approach which we also know from the SEM literature (e.g., Anderson & Gerbing, 1988). However, in contrast to SEM with latent variables, the first step conducts CCA, and subsequently, in the second step, the emergent variables are embedded in a structural model.

To collect the data, midsize firms in Spain were surveyed, using questionnaire items measured on a 5-point scale, to which the researchers received 100 valid responses. For the sake of ease and in line with Rhemtulla, Brosseau-Liard, and Savalei (2012) (see also Schubert, Henseler, & Dijkstra, 2018c), we treated the variables as continuous. Table 1 lists the 16 observable variables and gives the wording of the respective questionnaire items. Further, it shows the concepts that are intended to be constructed based on the corresponding blocks of observable variables. For more details on the data collection and the concepts, we refer the reader to the original study.

Based on the theory they derived, Benitez, Ray, et al. (2018) specified the composite model as depicted in Fig. 3.⁷ As illustrated, it assumed that each concept is composed of four observable variables. Additionally, all constructs were allowed to correlate freely, i. e., no constraints were imposed on the correlation matrix of the emergent variables. Finally, a CCA was conducted to examine whether their specified model was consistent with the collected data.

Considering the identification of the specified model, we note that no emergent variable was isolated in the model. Moreover, the weights were scaled to obtain emergent variables with unit variance. To ensure that the signs of the weights were uniquely determined, the first observable variable of each emergent variable, namely COMP1, CONN1, MOD1, and PSF1, was used as dominant indicator, i. e., the weights were chosen in a way that ensured a positive correlation between the observable variables and their respective emergent variables. Consequently, the model parameters can, in theory, be uniquely retrieved from the variance–covariance matrix of the observable variables.

To estimate the specified model, we followed the original study of Benitez, Ray, et al. (2018) and applied the iterative PLS algorithm using Mode A to obtain the weight and the construct correlation estimates. Mode A was deliberately chosen for the estimation because of the

⁶ We thank José Benitez and Gautam Ray for providing their data for the purpose of this demonstration.

⁷ It is noted that the covariances among the indicators of each block are concealed to preserve clarity.

Table 1
Items used in the CCA. Taken from Benitez, Ray, et al. (2018, p. A2).

Concept	Variable Name	Item Wording
IT compatibility	COMP1	Software applications can be easily transported and used across multiple platforms.
	COMP2	Our firm provides multiple interfaces or entry points (e.g., web access) for external end users.
	COMP3	Our firm establishes corporate rules and standards for hardware and operating systems to ensure platform compatibility.
	COMP4	Data captured in one part of our organization are immediately available to everyone in the firm.
IT connectivity	CONN1	Our organization has electronic links and connections throughout the entire firm.
	CONN2	Our firm is linked to business partners through electronic channels (e.g., websites, e-mail, wireless devices, electronic data interchange).
	CONN3	All remote, branch, and mobile offices are connected to the central office.
Modularity	CONN4	There are very few identifiable communications bottlenecks within our firm.
	MOD1	Our firm possesses a great speed in developing new business applications or modifying existing applications.
	MOD2	Our corporate database is able to communicate in several different protocols.
	MOD3	Reusable software modules are widely used in new systems development.
IT personnel skills flexibility	MOD4	IT personnel use object-oriented and prepackaged modular tools to create software applications.
	PSF1	Our IT personnel have the ability to work effectively in cross-functional teams.
	PSF2	Our IT personnel are able to interpret business problems and develop appropriate technical solutions.
	PSF3	Our IT personnel are self-directed and proactive.
	PSF4	Our IT personnel are knowledgeable about the key success factors in our firm.

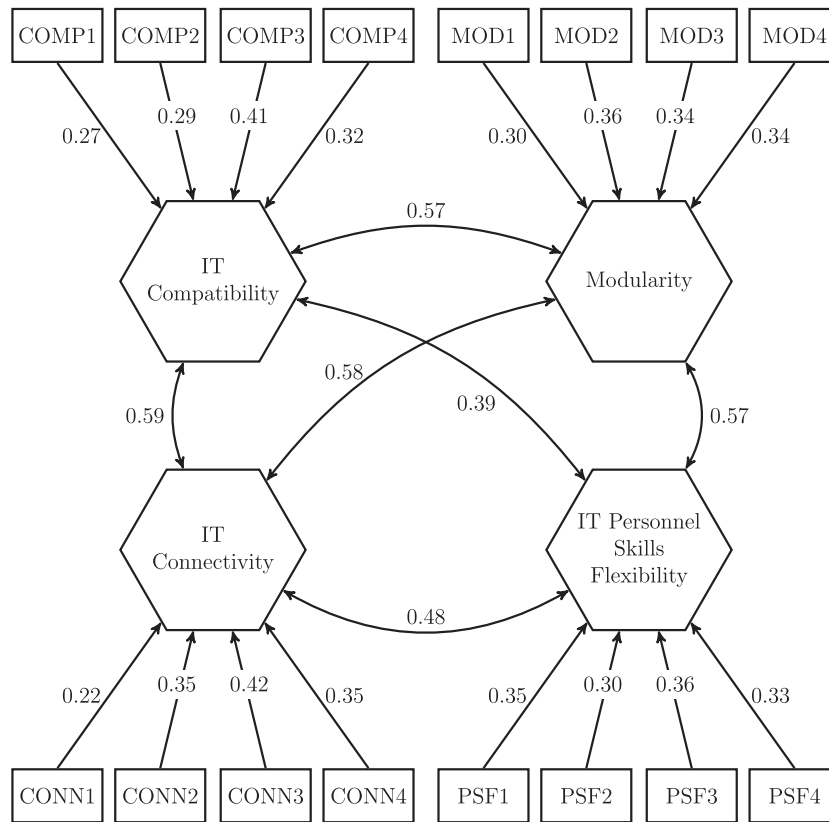


Fig. 3. Weight and covariance estimates obtained through CCA.

degree of collinearity among the indicators and the relatively small sample size.⁸ Additionally, we used the factorial scheme to calculate the inner weights in PLS. The whole analysis was conducted in the statistical programming environment R (R Core Team, 2020) using the cSEM package (Rademaker & Schubert, 2020).

We report the obtained weight estimates and the correlations between the observable variables and their construct, i.e., the loadings, in

⁸ Intra-block correlations range from 0.08 to 0.62. Moreover, if PLS Mode B was used, the variance inflation factors would range from 1.73 to 3.08 and some of the obtained weights would show negative signs and not differ significantly from 0, although they are expected to be positive and significant.

Table 2. Since there are no closed-form standard errors of the PLS parameter estimates, the statistical inference is based on bootstrap. In doing so, the corresponding 95% percentile confidence intervals based on 999 bootstrap runs are used. Moreover, the correlations among the emergent variables range from 0.39 to 0.59, and none of their 95% percentile confidence intervals covered the 0.

Assessing the composite model involves two steps, i.e., the assessment of the overall model fit and the assessment of each emergent variable separately. To assess the overall model fit, we used the bootstrap-based test for exact overall model fit. The results displayed in **Table 3** show that the values of the discrepancy measures, i.e., geodesic distance (d_G), SRMR, and squared Euclidean distance (d_t), are below the corresponding critical value, namely the 95% quantile of the

Table 2
Weight and loading estimates including their confidence intervals.

Concept	Variable	Weight	95% CI		Loading	95% CI	
IT compatibility	COMP1	0.2705	0.1827	0.3666	0.6670	0.4911	0.7973
	COMP2	0.2869	0.1934	0.3636	0.8071	0.6472	0.8847
	COMP3	0.4056	0.3331	0.5094	0.8380	0.7404	0.8936
	COMP4	0.3230	0.2441	0.4176	0.7680	0.6201	0.8542
IT connectivity	CONN1	0.2236	0.0923	0.3325	0.5311	0.2421	0.7315
	CONN2	0.3548	0.2630	0.4317	0.7984	0.6721	0.8733
	CONN3	0.4227	0.3557	0.5105	0.8279	0.7340	0.8930
	CONN4	0.3475	0.2512	0.4494	0.7139	0.5457	0.8317
Modularity	MOD1	0.2991	0.2138	0.3777	0.7042	0.5631	0.8065
	MOD2	0.3551	0.2931	0.4360	0.7933	0.6824	0.8645
	MOD3	0.3444	0.2801	0.4062	0.8218	0.7258	0.8831
	MOD4	0.3435	0.2660	0.4462	0.6541	0.4674	0.7792
IT personnel skills flexibility	PSF1	0.3487	0.2484	0.5060	0.7605	0.6312	0.8685
	PSF2	0.2974	0.1357	0.3941	0.7243	0.5349	0.8294
	PSF3	0.3631	0.2640	0.4736	0.7295	0.5508	0.8367
	PSF4	0.3317	0.2040	0.4351	0.7674	0.6287	0.8600

Table 3
Overall model fit assessment.

Discrepancy Measure	Test Statistic	Critical Value ($\alpha = 5\%$)
d_G	0.3036	0.3370
SRMR	0.0701	0.0753
d_L	0.6677	0.7720

corresponding reference distribution. The results lead us to conclude that the specified model adequately fits the collected data, i.e., the proposed model captures the available information in the data acceptably.⁹

As a second step of model assessment, each emergent variable is considered separately, i.e., we assessed the model locally. In line with Benitez, Ray, et al. (2018)'s expectations and as shown in Table 2, all observable variables significantly contribute to their emergent variable, i.e., the 95% percentile confidence intervals of the estimated weights do not cover the zero. Similarly, all correlations between the observable variables and their construct, as well as the correlations between the emergent variables are positive and significantly different from zero, i.e., the 95% percentile confidence intervals do not cover the zero.

Considering the results of the model assessment, we found no empirical evidence against the specified model, and thus the postulated theory of Benitez, Ray, et al. (2018) cannot be falsified. Notably, this lack of disconfirmation does not automatically imply a confirmation of the theory. As with all empirical studies, replicating the study is crucial to obtain more confidence in the model.

4. Discussion

The empirical example given above serves as showcase to illustrate how a statistical method, namely CCA, is applied. Hence, it should be obvious that an empirical study generally involves much more than

⁹In case that the test for overall model fit would have been significant, it indicates that the data comprises more information than captured by the specified model (Jöreskog, 1969). As a consequence, researchers should investigate potential sources for the misfits. In doing so, they can follow existing guidelines known from SEM, e.g., Kline (2015, Chapter 12), and inspect the variance–covariance of the residuals, i.e., the difference between the sample variance–covariance matrix and the model-implied counterpart. Moreover, they can consult fit indices to investigate whether the approximate fit of the model is still acceptable. As CCA was developed only recently, it is up to future research to propose more sophisticated strategies to identify sources of misfits in composite models, e.g., the evaluation of local fit, which has been recently proposed in the context of latent variable models (Thoemmes, Rosseel, & Textor, 2018).

simply undertaking the CCA steps presented in this paper. For example, as is common in quantitative research, researchers need to develop a theory and derive testable hypotheses prior to the data analysis. Moreover, researchers must decide about the way in which to collect the data for their studies and address associated issues such as common method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Although several statistical procedures to assess common method variance have been proposed in the context of common factor models (see, e.g., Williams, Hartman, & Cavazotte, 2010; Fuller, Simmering, Atinc, Atinc, & Babin, 2016), it is currently not clear how to address common method variance in the context of composite models. In general, researchers are advised to take various forms of uncertainty into account (Rigdon, Sarstedt, & Becker, 2020). Further, researchers should investigate a priori whether the size of the collected sample is sufficiently large to ensure that the employed statistical tests have sufficient power, e.g., by conducting Monte Carlo simulations (see e.g., Wolf, Harrington, Clark, & Miller, 2013). However, in the context of composite models, to date no guidelines have been proposed for setting up such a simulation. Finally, we need to assess whether the employed statistical methods and their associated assumptions match theory's implications and the characteristics of the collected data.

Properties of statistical methods including estimators and statistical tests are derived under assumptions that are often not met in empirical research, such as linearity or i.i.d. random variables. Moreover, often properties such as consistency and convergence in distribution refer to asymptotic behaviors, i.e., when the sample size converges to infinity, which is obviously never the case in empirical research. Hence, it is important to be aware of which assumptions are met and what their implied limitations for empirical research are. It is also important to know what and how serious the consequences are if these assumptions are violated.

This highlights the importance of methodological research that explores the limitations of statistical methods, such as consistent estimators' behavior in finite samples, and develops new methods that relax the original assumptions and take issues encountered in empirical research into account. As CCA has been introduced only recently, many of the methodological enhancements proposed in the context of SEM, such as equivalence testing (Yuan, Chan, Marcoulides, & Bentler, 2015), are not available yet. This endorses the model of the null hypothesis rather than rejecting it. It also endorses the test of close fit based on the SRMR (Maydeu-Olivares, Shi, & Rosseel, 2017), which works well even for smaller sample sizes. Similarly, current literature lacks proper guidelines on how to deal with sources of model misfit in the context of composite models. However, as CCA is still in its infancy, we are convinced as an analytic technique it will experience similar enhancements as CFA and SEM, and thus overcome limitations currently faced.

5. Conclusion

Constructs are the core building blocks of statistical models in CSA. Two types of constructs can be distinguished: latent variables and emergent variables. Latent variables are the common cause underlying a set of indicators. They are particularly useful to model theoretical concepts of behavioral research, such as traits and attitudes. Typically, they are extracted as common factors. The preferred method to empirically assess them is CFA. In contrast, emergent variables are typically defined by a set of observable variables, i. e., they are composites of other variables, which are particularly useful to model human-made artifacts. For a long time, there has been no method to empirically assess them.

In 2014, Jörg Henseler and Theo K. Dijkstra developed and proposed a statistical method that assesses the goodness of fit of models of interrelated composites, and – making use of the inventor’s privilege to name the invention – they labeled it ‘confirmatory composite analysis’ (Henseler et al., 2014, see also the authors’ note in that paper). The full presentation of the method appeared in Schuberth et al. (2018b), and it is now being introduced to business research by means of the current paper, which explains all of CCA’s main steps. Not only does the name CCA emphasize the analogy to CFA, with which CCA shares everything excepting the specified model, but it also clearly designates what CCA actually does: it enables the *analysis of composite* models, and it is *confirmatory* in nature because it provides evidence about whether the collected data is consistent with a researcher’s specified composite model. Thus, the name explicitly articulates a central aspect of the meaning of this analytic approach.

In their recent JBR paper, Hair et al. (2020) used the term ‘confirmatory composite analysis’ for something else: the measurement model evaluation step of PLS-SEM. In doing so, they appear to associate CCA with nothing more than a body of existing rules of thumb and cookbook procedures, which have for some time been associated with PLS path modeling, in particular. Briefly, according to Hair et al. (2020), CCA is little more than a new name for an old recipe.¹⁰ This is unfortunate, because it could unnecessarily confuse business researchers.

In statistical modeling, it has become a good practice to distinguish between the statistical model and its unknown parameters on the one hand, and the employed estimator as a function of the data on the other hand (Lehmann & Casella, 2003, p. 4). Researchers are thus confronted with two problems in decision making. First, they need to specify the statistical model, which includes for example the decision about the number of free and fixed parameters. Second, they must choose an appropriate estimator to obtain estimates for the free model parameters. Preferably, estimators should have certain desired (asymptotic) properties, such as being unbiased, consistent, and efficient.¹¹ In order to establish such properties, it is necessary that certain assumptions, which are estimator specific, are fulfilled. If these assumptions are violated, an estimator loses its properties. Consequently, its estimates turn out to be biased and inconsistent, as do entities that rely on the estimates, such as their standard errors. Moreover, statistical inference leads to wrong conclusions, which renders a study’s findings questionable. In particular, an estimator must suit the statistical model. For instance, it is well known that using estimators assuming a composite model such as PLS and GSCA to estimate the parameters of common factor models/reflective measurement models leads to biased estimates (Dijkstra, 1983; Dijkstra & Henseler, 2015; Hwang, Takane, & Jung, 2017), which can inflate Type-I and Type-II errors (Goodhue, Lewis, & Thompson, 2017; Henseler, 2012). The opposite constellation is equally problematic: using consistent PLS (PLSc, Dijkstra & Henseler, 2015),

which assumes a common factor model, to estimate the parameters of a composite model also leads to biased estimates (Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016).

In the light of this situation, our paper offers two pieces of advice. First, scholars should not confuse Hair et al.’s recommendations with the CCA defined in the literature. CCA is *not* the measurement model evaluation step of PLS-SEM, but an innovative set of procedures for specifying and assessing composite models. Second, analysts should always assess a model with a method that has been designed for precisely such a model. Concretely, CFA is the preferred method to assess common factor models (reflective measurement models), and CCA is the method of choice to assess composite models.

References

- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike. Springer series in statistics (perspectives in statistics)* (pp. 199–213). New York: Springer. https://doi.org/10.1007/978-1-4612-1694-0_15.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94. <https://doi.org/10.1007/BF02723327>.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*, 815–824. <https://doi.org/10.1016/j.paid.2006.09.018>.
- Benítez-Ávila, C., Hartmann, A., Dewulf, G., & Henseler, J. (2018). Interplay of relational and contractual governance in public-private partnerships: The mediating role of relational norms, trust and partners’ contribution. *International Journal of Project Management*, *36*, 429–443. <https://doi.org/10.1016/j.ijproman.2017.12.005>.
- Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*, *2*, 103168. <https://doi.org/10.1016/j.im.2019.05.003>.
- Benitez, J., Llorens, J., & Braojos, J. (2018). How information technology influences opportunity exploration and exploitation firm’s capabilities. *Information & Management*, *55*, 508–523. <https://doi.org/10.1016/j.im.2018.03.001>.
- Benitez, J., Ray, G., & Henseler, J. (2018). Impact of information technology infrastructure flexibility on mergers and acquisitions. *MIS Quarterly*, *42*, 25–43. <https://doi.org/10.25300/MISQ/2018/13245>.
- Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, *13*, 95–115. <https://doi.org/10.1214/aos/1176346579>.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, *6*, 25–53. <https://doi.org/10.1080/15366360802035497>.
- Bozdogan, H. (1987). Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370. <https://doi.org/10.1007/bf02294361>.
- Braojos, J., Benitez, J., & Llorens, J. (2019). How do social commerce-IT capabilities influence firm performance? Theory and empirical evidence. *Information & Management*, *56*, 155–171. <https://doi.org/10.1016/j.im.2018.04.006>.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Thousand Oaks, CA: Sage.
- Cegarra-Navarro, J. G., Ruiz, F. J. A., Martínez-Caro, E., & García-Pérez, A. (in press). Turning heterogeneity into improved research outputs in international R&D teams. *Journal of Business Research* (p. in print). doi:10.1016/j.jbusres.2019.05.023.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. New York: Springer.
- Cegarra-Navarro, J. G., Papa, A., García-Pérez, A., & Fiano, F. (2019). An open-minded strategy towards eco-innovation: A key to sustainable growth in a global enterprise. *Technological Forecasting and Social Change*, *148*, 119727. <https://doi.org/10.1016/j.techfore.2019.119727>.
- Chin, W. W. (2010). How to write up and report PLS analyses. In V. Esposito Vinzi, W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 655–690). Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-32827-8_29.
- Choi, J. Y., & Hwang, H. (2020). Bayesian generalized structured component analysis. *British Journal of Mathematical and Statistical Psychology*, *73*, 347–373. <https://doi.org/10.1111/bmsp.12166>.
- Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review*, *2*, 274–285. <https://doi.org/10.1177/1754073910361978>.
- Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement*, *14*, 183–196. <https://doi.org/10.1177/014662169001400207>.
- Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*, 174–184. <https://doi.org/10.1037/0033-2909.114.1.174>.
- Dijkstra, T. K. (1983). Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics*, *22*, 67–90. [¹⁰ We thank an anonymous reviewer for providing this perspective.](https://doi.org/10.1016/0304-</p>
</div>
<div data-bbox=)

¹¹ For a detailed explanation of these behaviors, we refer to Lehmann and Casella (2003).

- 4076(83)90094-5.
- Dijkstra, T. K. (2015). *All-inclusive versus single block composites*. doi:10.13140/RG.2.1.2917.8082 working paper.
- Dijkstra, T. K. (2017). A perfect match between a model and a mode. In H. Latan, & R. Noonan (Eds.), *Partial least squares path modeling* (pp. 55–80). Cham: Springer. https://doi.org/10.1007/978-3-319-64069-3_4.
- Dijkstra, T. K., & Henseler, J. (2011). Linear indices in nonlinear structural equation models: Best fitting proper indices and other composites. *Quality & Quantity*, 45, 1505–1518. <https://doi.org/10.1007/s11135-010-9359-z>.
- Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, 39, 297–316. <https://doi.org/10.25300/MISQ/2015/39.2.02>.
- Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, 4, 144–192. <https://doi.org/10.1177/109442810142004>.
- Felipe, C. M., Roldán, J. L., & Leal-Rodríguez, A. L. (2016). An explanatory and predictive model for organizational agility. *Journal of Business Research*, 69, 4624–4631. <https://doi.org/10.1016/j.jbusres.2016.04.014>.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- Foltean, F. S., Trif, S. M., & Tuleu, D. L. (2019). Customer relationship management capabilities and social media technology use: Consequences on firm performance. *Journal of Business Research*, 104, 563–575. <https://doi.org/10.1016/j.jbusres.2018.10.047>.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39–50. <https://doi.org/10.2307/3151312>.
- Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of Business Research*, 69, 3192–3198. <https://doi.org/10.1016/j.jbusres.2015.12.008>.
- Goodhue, D. L., Lewis, W., & Thompson, R. L. (2017). A multicollinearity and measurement error statistical blind spot: Correcting for excessive false positives in regression and PLS. *MIS Quarterly*, 41, 667–684. <https://doi.org/10.25300/MISQ/2017/41.3.01>.
- Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research*, 109, 101–110. <https://doi.org/10.1016/j.jbusres.2019.11.069>.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 18, 139–152. <https://doi.org/10.2753/MTP1069-6679190202>.
- 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, 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>.
- Henseler, J. (2012). Why generalized structured component analysis is not universally preferable to structural equation modeling. *Journal of the Academy of Marketing Science*, 40, 402–413. <https://doi.org/10.1007/s11747-011-0298-6>.
- Henseler, J. (2015a). Confirmatory composite analysis. In *2nd International symposium on partial least squares path modeling*. Seville, Spain. June 16–19.
- Henseler, J. (2015b). Confirmatory composite analysis. In *5th modern modeling methods conference*. Storrs, CT, U.S.A. May 19–20.
- Henseler, J. (2015c). Is the whole more than the sum of its parts? On the interplay of marketing and design research. Enschede: University of Twente. <http://purl.utwente.nl/publications/95770>.
- Henseler, J. (2015d). Towards a framework for confirmatory empirical design research. In *5th CIM community workshop*. Enschede, the Netherlands. September 1–2.
- Henseler, J. (2017). Bridging design and behavioral research with variance-based structural equation modeling. *Journal of Advertising*, 46, 178–192. <https://doi.org/10.1080/00913367.2017.1281780>.
- 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, 182–209. <https://doi.org/10.1177/1094428114526928>.
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, 116, 2–20. <https://doi.org/10.1108/IMDS-09-2015-0382>.
- Hernández-Perlines, F., Moreno-García, J., & Yáñez-Araque, B. (2016). Training and business performance: The mediating role of absorptive capacities. *SpringerPlus*, 5, 2074. <https://doi.org/10.1186/s40064-016-3752-6>.
- Huang, P.-H. (2017). Asymptotics of AIC, BIC, and RMSEA for model selection in structural equation modeling. *Psychometrika*, 82, 407–426. <https://doi.org/10.1007/s11336-017-9572-y>.
- Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>.
- Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, 69, 81–99. <https://doi.org/10.1007/BF02295841>.
- Hwang, H., Takane, Y., & Jung, K. (2017). Generalized structured component analysis with uniqueness terms for accommodating measurement error. *Frontiers in Psychology*, 8, 2137. <https://doi.org/10.3389/fpsyg.2017.02137>.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202. <https://doi.org/10.1007/bf02289343>.
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58, 433–451. <https://doi.org/10.1093/biomet/58.3.433>.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.
- Lehmann, E. L., & Casella, G. (2003). *Theory of point estimation*. New York: Springer.
- Marcoulides, G. A., & Chin, W. W. (2013). You write, but others read: Common methodological misunderstandings in PLS and related methods. In H. Abdi, W. W. Chin, V. Esposito Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New perspectives in partial least squares and related methods* (pp. 31–64). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-8283-3_2.
- Martelo-Landroguez, S., Cegarra Navarro, J.-G., & Cepeda-Carrión, G. (2019). Uncontrolled counter-knowledge: Its effects on knowledge management corridors. *Knowledge Management Research & Practice*, 17, 203–212. <https://doi.org/10.1080/14778238.2019.1599497>.
- Martin, E. S., & Quintana, F. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, 16, 99–106.
- Mason, R., & Brown, W. G. (1975). Multicollinearity problems and ridge regression in sociological models. *Social Science Research*, 4, 135–149. [https://doi.org/10.1016/0049-089x\(75\)90008-3](https://doi.org/10.1016/0049-089x(75)90008-3).
- Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2017). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 389–402. <https://doi.org/10.1080/10705511.2017.1389611>.
- McQuarrie, A. D. R., & Tsai, C.-L. (1998). *Regression and time series model selection*. World Scientific. <https://doi.org/10.1142/3573>.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430–445. <https://doi.org/10.1037/0033-2909.105.3.430>.
- Nuzzo, R. (2015). How scientists fool themselves – and how they can stop. *Nature*, 526, 182–185. <https://doi.org/10.1038/526182a>.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559–572. <https://doi.org/10.1080/14786440109462720>.
- Petter, S., Straub, D. W., & Rai, A. (2007). Specifying formative constructs in Information Systems research. *MIS Quarterly*, 31, 623–656. <https://doi.org/10.2307/25148814>.
- Pittino, D., Martínez, A. B., Chirico, F., & Galván, R. S. (2018). Psychological ownership, knowledge sharing and entrepreneurial orientation in family firms: The moderating role of governance heterogeneity. *Journal of Business Research*, 84, 312–326. <https://doi.org/10.1016/j.jbusres.2017.08.014>.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>.
- Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, 17, 1–14. <https://doi.org/10.1037/a0026804>.
- Rademaker, M., & Schuberth, F. (2020). *cSEM: Composite-based structural equation modeling*. <https://github.com/M-E-Rademaker/cSEMPackageversion.0.1.0.9000>.
- Rasoolimanesh, S. M., Md Noor, S., Schuberth, F., & Jaafar, M. (2019). Investigating the effects of tourist engagement on satisfaction and loyalty. *The Service Industries Journal*, 39, 559–574. <https://doi.org/10.1080/02642069.2019.1570152>.
- R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Reise, S. P. (1999). Measurement issues viewed through the eyes of IRT. In S. E. Embretson, & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 219–241). New York & London: Psychology Press.
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17, 354–373. <https://doi.org/10.1037/a0029315>.
- Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, 25, 30–45. <https://doi.org/10.1037/met0000220>.
- Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, 45, 341–358. <https://doi.org/10.1016/j.lrp.2012.09.010>.
- Rigdon, E. E., Sarstedt, M., & Becker, J.-M. (2020). Quantify uncertainty in behavioral research. *Nature Human Behaviour*, 4, 329–331. <https://doi.org/10.1038/s41562-019-0806-0>.
- Roldán, J. L., Real, J. C., & Sánchez Ceballos, S. (2018). Antecedents and consequences of knowledge management performance: The role of IT infrastructure. *Intangible Capital*, 14, 518–535. <https://doi.org/10.3926/ic.1074>.
- Rueda, L., Benitez, J., & Braojos, J. (2017). From traditional education technologies to student satisfaction in management education: A theory of the role of social media applications. *Information & Management*, 54, 1059–1071. <https://doi.org/10.1016/j.im.2017.06.002>.
- Ruiz-Palomo, D., Diéguez-Soto, J., Duréndez, A., & Santos, J. A. C. (2019). Family management and firm performance in family SMEs: The mediating roles of management control systems and technological innovation. *Sustainability*, 11, 3805. <https://doi.org/10.3390/su11143805>.
- Sajtos, L., & Magyar, B. (2016). Auxiliary theories as translation mechanisms for measurement model specification. *Journal of Business Research*, 69, 3186–3191. <https://doi.org/10.1016/j.jbusres.2015.12.007>.
- Sanchez-Franco, M. J., Cepeda-Carrión, G., & Roldán, J. L. (2019). Understanding relationship quality in hospitality services: A study based on text analytics and partial least squares. *Internet Research*, 29, 478–503. <https://doi.org/10.1108/IntR-12-2017-0531>.
- Sánchez-Polo, M. T., Cegarra-Navarro, J.-G., Cillo, V., & Wensley, A. (2019). Overcoming knowledge barriers to health care through continuous learning. *Journal of Knowledge Management*, 23, 508–526. <https://doi.org/10.1108/JKM-10-2018-0636>.
- 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, 3998–4010. <https://doi.org/10.1016/j.jbusres.2016.06.007>.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23–74.
- Schuberth, F., Dijkstra, T.K., & Henseler, J. (2018a). Confirmatory composite analysis. In *Meeting of the SEM working group. Amsterdam, the Netherlands. March 15–16*.
- Schuberth, F., Rademaker, M., & Henseler, J. (in press). Estimating and assessing second-order constructs using PLS-PM: the case of composites of composites. *Industrial Management & Data Systems*.
- Schuberth, F., Henseler, J., & Dijkstra, T. K. (Henseler, & Dijkstra, 2018b). Confirmatory composite analysis. *Frontiers in Psychology*, 9, 2541. <https://doi.org/10.3389/fpsyg.2018.02541>.
- Schuberth, F. (in press). Confirmatory composite analysis using partial least squares: Setting the record straight. *Review of Managerial Science*. doi:10.1007/s11846-020-00405-0.
- Schuberth, F., Henseler, J., & Dijkstra, T. K. (Henseler, & Dijkstra, 2018c). Partial least squares path modeling using ordinal categorical indicators. *Quality & Quantity*, 52, 9–35. <https://doi.org/10.1007/s11135-016-0401-7>.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464. <https://doi.org/10.1214/aos/1176344136>.
- Thoemmes, F., Rosseel, Y., & Textor, J. (2018). Local fit evaluation of structural equation models using graphical criteria. *Psychological Methods*, 23, 27–41. <https://doi.org/10.1037/met0000147>.
- van Riel, A. C., Henseler, J., Kemény, I., & Sasovova, Z. (2017). Estimating hierarchical constructs using consistent partial least squares: The case of second-order composites of common factors. *Industrial Management & Data Systems*, 117, 459–477. <https://doi.org/10.1108/IMDS-07-2016-0286>.
- Vidaurre, D., van Gerven, M. A., Bielza, C., Larrañaga, P., & Heskes, T. (2013). Bayesian sparse partial least squares. *Neural Computation*, 25, 3318–3339. https://doi.org/10.1162/NECO_a_00524.
- Wei, S., Ke, W., Liu, H., & Wei, K. K. (2020). Supply chain information integration and firm performance: Are explorative and exploitative IT capabilities complementary or substitutive? *Decision Sciences*, 51, 464–499. <https://doi.org/10.1111/dec.12364>.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.). *Handbook of structural equation modeling* (pp. 209–231). New York: Guilford Press.
- Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods*, 13, 477–514. <https://doi.org/10.1177/1094428110366036>.
- Wold, H. O. A. (1973). Nonlinear iterative partial least squares (NIPALS) modelling. Some current developments. In P. R. Krishnaiah (Ed.), *Proceedings of the 3rd international symposium on multivariate analysis* (pp. 383–407). Dayton, OH.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73, 913–934.
- Yiu, H. L., Ngai, E. W. T., & Lei, C. F. (2020). Impact of service-dominant orientation on the innovation performance of technology firms: Roles of knowledge sharing and relationship learning. *Decision Sciences*, 51, 620–654. <https://doi.org/10.1111/dec.12408>.
- Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, 40, 115–148. https://doi.org/10.1207/s15327906mbr4001_5.
- Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2015). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 319–330. <https://doi.org/10.1080/10705511.2015.1065414>.

Jörg Henseler holds the Chair of Product-Market Relations in the Faculty of Engineering Technology at the University of Twente, the Netherlands, and he is Visiting Professor at NOVA Information Management School, Universidade Nova de Lisboa, Portugal. His broad-ranging research interests encompass empirical methods of Marketing and Design research as well as the management of design, products, services, and brands. He is co-inventor of consistent partial least squares (PLSc), the heterotrait-monotrait ratio of correlations (HTMT), and confirmatory composite analysis (CCA). He is a highly cited researcher according to Web of Science; his work has been published in *Computational Statistics and Data Analysis*, *European Journal of Information Systems*, *International Journal of Research in Marketing*, *Journal of the Academy of Marketing Science*, *Journal of Supply Chain Management*, *MIS Quarterly*, *Organizational Research Methods*, and *Structural Equation Modeling-A Multidisciplinary Journal*, among others. He chairs the Scientific Advisory Board of ADANCO, software for composite-based SEM (<http://www.composite-modeling.com>).

Florian Schuberth obtained his PhD in Econometrics in the Faculty of Business Management and Economics at the University of Würzburg, Germany. Currently, he is Assistant Professor in the Faculty of Engineering Technology at the University of Twente, the Netherlands. His main research interests are focused on SEM, in particular on composite-based estimators and their enhancement. His work has been published in various journals such as *Behaviormetrika*, *Information and Management*, *Internet Research*, *Journal of Business Research*, and *Quality & Quantity*.