Counterpoint: Representing Forged Concepts as Emergent Variables Using Composite-Based Structural Equation Modeling

Xi Yu

University of Twente

Sam Zaza

Middle Tennessee State University

Florian Schuberth

University of Twente

Jörg Henseler

University of Twente

Universidade Nova de Lisboa

This is the author accepted manuscript version of the paper published by ACM as: Yu, X., Zaza, S., Schuberth, F., & Henseler, J. (2021). Counterpoint: Representing Forged Concepts as Emergent Variables Using Composite-Based Structural Equation Modeling. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 52(SI - Special Issue on Composite-based Structural Equation Modeling), 114-130.
https://doi.org/10.1145/3505639.3505647

Representing Forged Concepts as Emergent Variables Using Composite-based Structural Equation Modeling

Xi Yu   Sam Zaza   Florian Schuberth   Jörg Henseler

December 30, 2021

Xi Yu, Faculty of Engineering Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
E-mail: x.yu-1@utwente.nl
Sam Zaza, Information Systems and Analytics, Jones College of Business, Middle Tennessee State University, 1301 E Main St, Murfreesboro, TN 37132, USA
E-mail: sam.zaza@mtsu.edu
Florian Schuberth, Faculty of Engineering Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
E-mail: f.schuberth@utwente.nl
Jörg Henseler, Faculty of Engineering Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands and Nova Information Management School, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal
E-mail: j.henseler@utwente.nl

Abstract

Studying and modeling theoretical concepts is a cornerstone activity in Information Systems (IS) research. Researchers have been familiar with one type of theoretical concept, namely behavioral concepts, which are assumed to exist in nature and measured by a set of observable variables. In this paper, we present a second type of theoretical concept, namely forged concepts, which are designed and assumed to emerge within their environment. While behavioral concepts are operationalized as latent variables, forged concepts are better specified as emergent variables. Additionally, we propose composite-based structural equation modeling (SEM) as a subtype of SEM that is eminently suitable to analyze models containing emergent variables. We shed light on the composite-based SEM steps: model specification, model identification, model estimation, and model assessment. Then, we present an illustrative example from the domain of IS research to demonstrate these four steps and to show how modeling with emergent variables proceeds.

Keywords Composite-based structural equation modeling, emergent variables, composite model, forged concept, behavioral concept

1 Introduction

Theoretical concepts – formalized ideas of unobserved properties or attributes – are one of the key elements of a researcher's theory (Whetten, 1989). To model and assess researchers' theories comprising these concepts, structural equation modeling (SEM) is used in many disciplines such as Information Systems (Petter et al., 2007; Rutner et al., 2008). In SEM, theoretical concepts need to be operationalized, i.e., to be modeled as variables in a statistical model. Operationalizing theoretical concepts in SEM has gone through different waves of evolution.

In the first evolutionary wave, it was proposed to operationalize theoretical concepts as latent variables in a reflective measurement model. This idea was grounded in the true score theory (Novick, 1966), which dates back to Edgeworth (1888) and is supported by various theoretical frameworks, such as the holistic construal of organizational research (Bagozzi & Phillips, 1982). Applying the reflective measurement model, it is assumed that the observable variables are measurement error-prone manifestations of the theoretical concept. To date, the reflective measurement model is the dominant way of modeling theoretical concepts, and widely acknowledged in the IS literature. For instance, theoretical concepts such as Ease of Use (Brown & Venkatesh, 2005; Gefen et al., 2003; Karimi et al., 2004; Lewis et al., 2003; Van der Heijden, 2004; Wixom & Todd, 2005), Turnover Intention (Ahuja et al., 2007; Allen et al., 2009; Armstrong et al., 2018; Joseph et al., 2007; Moquin et al., 2019), and Role Overload (Ho et al., 2003) have been specified as latent variables incorporated in a reflective measurement model.

Recognizing that not all observable variables measuring a theoretical concept have to be measurement error-prone manifestations of it prompted a second evolutionary wave of theoretical concept operationalization, and then causal-formative measurement model arose (e.g., Bollen, 1984; Bollen & Lennox, 1991). The causal-formative measurement model originated from the idea of *cause indicators*, i.e., observable variables that affect the latent variable (Blalock, 1964; Jöreskog & Goldberger, 1975). It assumes that the observable variables are causal antecedents of theoretical concepts. Through its elaboration in various disciplines such as marketing (Diamantopoulos & Winklhofer, 2001) and consumer research (Jarvis et al., 2003), the causal-formative measurement model experienced an upsurge in the last two decades. In the

IS discipline, Petter et al. (2007) proposed the use of formative construct and causal-formative measurement model as an alternative way of concept operationalization. Their paper, which reached 3,270 citations by September 2021 (Google scholar), provides IS scholars with guidance on adequately distinguishing between reflective and causal-formative measurement models. Following their guidelines, theoretical concepts such as Social Influence (Johnston & Warkentin, 2010) and Third-party Assurance (Dimoka et al., 2012) have been specified as latent variables through a causal-formative measurement model.

Recently, a third evolutionary wave was triggered due to a new type of conceptualization and operationalization. While latent variables were introduced to model the unobserved properties of *social* units or entities (see Bagozzi & Phillips, 1982, p. 465), i.e., organisms, IS research also deals with the properties of *technical* units or entities, i.e., IT artifacts. These require a different type of conceptualization and operationalization. To emphasize that such concepts are made or designed, they were dubbed 'forged concepts' (Henseler, 2021). Instead of employing a latent variable to illustrate such concepts in the statistical model, it was proposed to use an emergent variable, i.e., a linear combination of variables, incorporated in the composite model (Henseler, 2015b, 2017; Schuberth et al., 2018). This idea was inspired by research that proposed employing composites as a summary of the effects in SEM (Bollen & Diamantopoulos, 2017a; Grace & Bollen, 2008; Heise, 1972). However, not all composites are emergent variables. While both are weighted linear combinations of variables, composite variables do not have to have conceptual unity (Bollen & Diamantopoulos, 2017a), whereas emergent variables are composite variables that have conceptual unity. The composite's status changed with the development of the composite model (Cho & Choi, 2020; Dijkstra, 2013, 2017; Henseler et al., 2014). Although the composite model also contains a composite at its core, namely the emergent variable, it additionally imposes restrictions on the observable variables' variance-covariance matrix, which can be exploited in statistical testing to empirically falsify a researcher's theory. Previous research that was supposed to model forged concepts, ignored this (see, e.g., Fornell & Bookstein, 1982). Moreover, modeling theoretical concepts by means of the composite model has proven useful in many fields, such as management research (Law & Wong, 1999), advertising (Henseler, 2017), tourism research (Müller et al., 2018; Rasoolimanesh et al., 2019), business research (Henseler & Schuberth, 2020b), and ecology (Grace & Bollen, 2008).

In the IS literature, Henseler et al. (2016) and Benitez et al. (2020) sketched this idea in the context of partial least squares path modeling (PLS-PM, Wold, 1975), an estimator that consistently estimates composite models (Dijkstra, 2017). However, modeling forged concepts is not only relevant in the context of PLS-PM, but also in other contexts. Hence, the necessity of expanding SEM's applicability to study forged concepts becomes evident. To address this issue, our study aims to shed light on forged concepts, their connection to composite-based SEM, and their operationalization as emergent variables. By doing so, we present the notion of forged concepts as a second type of theoretical concept and distinguish it from behavioral concepts that are typically studied in SEM.

The paper is structured as follows. In Section 2, we distinguish between behavioral and forged concepts and show the different ways of representing them in SEM. In Section 3, we explain composite-based SEM and present its successive steps. In Section 4, we demonstrate how composite-based SEM can be applied to a model with emergent variables through an illustrative example. Finally, we conclude the paper in Section 5 with a discussion and an outlook on future research.

2 Review of Theoretical Concepts

2.1 Types of Theoretical Concepts

At least two types of theoretical concepts can be distinguished in SEM, namely behavioral concepts and forged concepts (Benitez et al., 2020; Henseler, 2017; Henseler & Schuberth, 2020b; Hubona et al., 2021; Schuberth et al., 2018). In the following, we argue that the type of theoretical concept determines how a concept should be operationalized.

Following a scientific realist perspective, behavioral concepts are ontological entities assumed to exist in nature (Borsboom, 2008; Borsboom et al., 2003). Examples from IS research include Customer Behavior (Venkatesh et al., 2012; Zhang & Venkatesh, 2017), User Attitudes (Thatcher et al., 2018), Feelings (Armstrong et al., 2015), and Perceptions (Davis, 1989). To represent behavioral concepts in the structural model, latent variables are eminently suitable. Since the behavioral concept is assumed to be the common cause underlying a set of observable variables (Reichenbach, 1956), the observable variables are measurement error-prone manifestations of it; the reflective measurement model is suited to capture these characteristics. In contrast, if the behavioral concept is assumed to exist in nature and is caused by a set of observable variables, the causal-formative measurement model is a more obvious way of operationalizing such concepts (e.g., Bollen & Diamantopoulos, 2017a, 2017b; Hardin, 2017).

Besides behavioral concepts, forged concepts have been identified as another type of theoretical concept (Benitez et al., 2020; Henseler, 2017, 2021; Müller et al., 2018; Schuberth et al., 2018). Forged concepts emerge from the elements within their environment, as human constructions rather than naturally occurring phenomena (Henseler, 2017; Henseler & Schuberth, 2020a). Consequently, they are context-specific and not existent per se in nature. In the literature, this type of theoretical concept is also known as an *artifact* (Benitez et al., 2020; Henseler, 2017; Hubona et al., 2021; Müller et al., 2018; Schuberth et al., 2018). In the IS literature, there are some differences between artifacts and forged concepts. Artifacts are generally regarded as "things that have been, or can be transformed into, a material existence as artificially made objects (e.g., a model) or processes (e.g., method, software)" (Henseler, 2021, p. 31). However, forged concepts conceptualize attributes instead of objects. For example, distributed computing is an object and thus can be regarded as an artifact, while the distributivity of computing is an attribute that can be thought of as a forged concept. Imaginably, an artifact is possible to possess multiple forged concepts. Examples of forged concepts from IS research are Information Quality, System Quality, and Service Quality (Xu et al., 2013). Whereas Completeness, Accuracy, Format, and Currency serve as attributes in composing Information Quality, Reliability, Flexibility, Accessibility, and Timeliness define System Quality. Similarly, Tangibles, Responsiveness, Empathy, Service Reliability, and Assurance define Service Quality. In all of these examples, the dimensions do not cause the theoretical concepts; rather, they make them up.

2.2 Ways to Operationalize Theoretical Concepts

Petter et al. (2007) observed that reflective measurement models are too rigid in some situations, thus should not be used in modeling all theoretical concepts. Specifically, they argue that latent variables cannot only be modeled in reflective measurement models, but also causal-formative measurement models, giving the IS research community more freedom to model their phenomena than they could do before. We take it a step further and argue that researchers have more freedom in operationalizing their theoretical concepts than simply having to choose a latent variable in the form of a reflective or causal-formative measurement model. They can also choose which type of representation to employ for their theoretical concepts. Besides latent variables, they can use emergent variables. Table 1 depicts the various ways of modeling theoretical concepts in a structural equation model. Following Grace and Bollen (2008), composites and thus emergent variables are depicted by hexagons to distinguish them from latent variables, which are usually depicted by ovals.

We find latent variables in the reflective and the causal-formative measurement model to operationalize behavioral concepts. In the reflective measurement model, a latent variable explains the covariance structure of the observable variables. As shown In Table 1, in the reflective measurement model, the arrows are initiated from the latent variable to observable variables. To capture the variance in the observable variables that cannot be explained by the latent variables, so-called unique factors, which are also latent variables, account for the remaining variance in the observable variables. In the classical reflective measurement model, these unique factors are assumed to be uncorrelated. Consequently, the observable variables are uncorrelated when controlled for the latent variable. The dominant statistical methods to assess reflective measurement models are confirmatory factor analysis (CFA) and factor-based structural equation modeling (SEM).

In contrast to the reflective measurement model, in the causal-formative measurement model, the relationship between the observable variables and the latent variable is reversed (Diamantopoulos et al., 2008). Consequently, the arrows point from the observable variables to the latent variable. Since the observable variables most likely do not explain all the latent variable's variation, a disturbance term accounts for the remaining variation in the construct. The dominant statistical method to assess causal-formative measurement models is factor-based SEM.

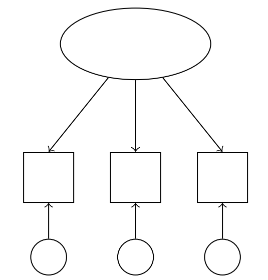
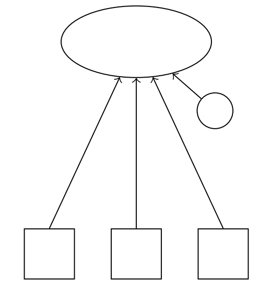
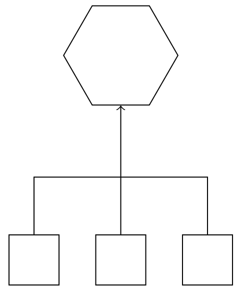
Considering forged concepts, measurement models, regardless of whether they are reflective or causal-formative, are hardly suitable for their operationalization. Both types of measurement models assume that 1) the theoretical concept exists in nature, and 2) there is a causal relationship between the theoretical concept and the observable variables. These assumptions ignore an important aspect of forged concepts, namely that they are human-made and do not otherwise exist in nature. Forged concepts have a definitorial relationship rather than a causal one with their observable variables. To capture these characteristics, the composite model is well suited to their operationalization (Henseler, 2015a; Henseler, 2017).

In the composite model, the theoretical concept is represented by an emergent variable, i.e., a linear combination of the observable variables, which can be regarded as a simplification of the forged concept's composition. As displayed on the right-hand side of Table 1, the emergent variable is at the core of the composite model. In contrast to the two measurement models that contain latent variables, in the composite model, it is assumed that the construct is fully composed of the observable variables. Hence, there is no error term on the construct level. The dominant statistical methods to assess composite models are confirmatory composite analysis (CCA) and composite-based SEM.

CCA is an innovative statistical method to specify and assess composite models, designed analogous to CFA, and includes the same steps: model specification, model identification, model estimation, and model assessment (Henseler & Schuberth, 2020b; Hubona et al., 2021; Schuberth, 2021; Schuberth et al., 2018). The difference between CCA and CFA is that CCA is designed to assess models with emergent variables, whereas CFA is applied to assess reflective measurement models containing latent variables. As originally proposed, in the first step of CCA, a model is specified which relates two or more emergent variables. However, models where one emergent variable correlates with other variables are also conceivable. To ensure that the model is identified, researchers need to determine the scale of the emergent variables and to ensure there is no isolated emergent variable in the model. The third step is to estimate the model to obtain the model parameter estimates. To assess the model, researchers should consider the overall model fit, the parameter estimates, including their significances, and other metrics as proposed by Henseler and Schuberth (2020b).

To conclude, while Petter et al. (2007, p. 624) "call both the constructs and measures either formative or reflective", we distinguish between latent and emergent variables, between behavioral and forged concepts, and between reflective measurement, causal-formative measurement, and composite models. In doing so, we argue that the nature of the theoretical concept should determine its operationalization. Specifically, we argue that a behavioral concept, i.e., a theo-

Table 1: Ways of operationalizing theoretical concepts in SEM

Type of Concept:	Behavioral Concept		Forged Concept
Type of Construct:	Latent Variable	Latent Variable	Emergent Variable
Type of Model:	Reflective Measurement Model	Causal-formative Measurement Model	Composite Model
			
Statistical Method:	CFA / SEM	SEM	CCA / SEM
Evolutionary Wave:	1st wave: 1980's	2nd wave: 2000's	3rd wave: 2020's

Note: Squares symbolize observable variables; ovals symbolize latent variables, hexagons symbolize emergent variables; circles symbolize disturbances; direct arrows represent causal relationships; and fork arrows represent definitorial relationships. For further details, see Appendix A.

retical concept assumed to exist, should be modeled as a latent variable. In contrast, a forged concept, i.e., a theoretical concept assumed to be human-made, should be modeled as an emergent variable incorporated in the composite model. Further, researchers should employ either the reflective or the causal-formative measurement depending on the assumption regarding the relationship between the observable variables and the behavioral concept.

3 Composite-Based Structural Equation Modeling

SEM is a powerful approach to studying the relationships between theoretical concepts. In general, it comprises the following four steps (Schumacker & Lomax, 2009, Chapter 4): model specification, model identification, model estimation, and model assessment. Notably, the estimator choice in the estimation step has a significant impact on the other three steps because it determines which models can be specified, and which identification rules need to be employed. Moreover, the employed estimator and its statistical properties determine the opportunities of model assessment.

SEM estimators can be determined in several ways (Henseler, 2021, Chapter 1). First, estimators can be distinguished with regard to the model they assume and estimate. While certain estimators, such as the full information maximum likelihood (ML) estimator proposed by Jöreskog (1970), have been specifically tailored for estimating latent variable models, other estimators, such as generalized structured component analysis (GSCA, Hwang & Takane, 2004), can be used to consistently estimate models containing solely emergent variables in the structural model. Moreover, certain estimators such as ML (Henseler, 2021; Schuberth, in press), consistent partial least squares (PLSc, Dijkstra & Henseler, 2015b; Rademaker et al., 2019), and integrated generalized structured component analysis (IGSCA, Hwang et al., 2021) can deal with models containing both latent and emergent variables.

Second, estimators can be differentiated according to the employed optimization criterion, i.e., a variance-based or covariance-based estimator. Variance-based estimators, such as PLS-PM (Wold, 1975), create linear combinations of observable variables as proxies for the constructs, such that a certain criterion of interrelatedness between these proxies is optimized. Subsequently, the model parameters are estimated based on these proxies. In contrast, to obtain the parameter estimates, covariance-based estimators minimize the discrepancy between the empirical and the model-implied variance-covariance matrix of the observable variables.

Finally, SEM estimators can be distinguished by full-information and limited information estimators (Lance et al., 1988). Full information estimators such as GSCA utilize intra- and inter-equation information to estimate the model parameters. In contrast, limited information estimators such as PLS-PM use only the intra-equation information, i.e., the information relevant to the equation under investigation, while the results of the other equations are regarded as being given.

Only recently, extant literature distinguished between two types of SEM, namely composite-based and factor-based SEM (e.g., Hair et al., 2017; Hwang et al., 2021; Rigdon, 2012; Rigdon, 2016; Sarstedt et al., 2016). Following Rigdon's definition, in factor-based SEM, theoretical concepts are represented by common factors, i.e., latent variables, while in composite-based SEM, the theoretical concepts are modeled as composites, i.e., emergent variables (Rigdon, 2012). This is in line with our distinction of the estimators with regard to the estimated model, i.e., common factor model (reflective measurement model) and composite model. Although we mainly agree with the current definitions of composite-based and factor-based SEM, they are unnecessarily limited. For instance, based on these definitions, PLSc would fall under factor-based SEM, although it builds on the composite model to estimate a common factor model. Therefore, it requires the identification rules of the composite model. Hence, it is more closely related to composite-based SEM than to factor-based SEM.

To address this shortcoming, we extend the current definition of composite-based SEM to include SEM that builds on composites in the estimation step. Consequently, composite-based and factor-based SEMs are shortcuts to describe the estimator type employed in the SEM's estimation step.

Composite-based SEM is a shortcut for SEM in which an estimator is employed that builds on composites to obtain the model parameters. Hence, it covers estimators that use composites in combination with a correction for attenuation to estimate latent variable models such as PLSc. Moreover, composite-based SEM comprises structural equation models containing emergent variables, i.e., composites do not only serve as auxiliary variables in the estimation, but also represent theoretical concepts in the form of emergent variables in the statistical model. Since composite-based SEM is a subtype of SEM, it follows the same four steps as SEM: (1) model specification, (2) model identification, (3) model estimation, and (4) model assessment. The four steps are elaborated in the following section.

3.1 Model Specification

In SEM, model specification is generally applied to transfer a researcher's theory into a statistical model (Bollen, 1989). The same is true for composite-based SEM. Consequently, a researcher needs to decide how a theoretical concept is represented in the statistical model, i.e., by means of a latent variable or an emergent variable. Following our recommendation from Section 2, forged concepts should be operationalized by the composite model, while concepts assumed to exist naturally, i.e., behavioral concepts, should be modeled via the reflective or the causal-formative measurement model. Considering the composite model comprising the emergent variable, a researcher needs to decide about the variables forming the emergent variable. In doing so, emergent variables formed by observable variables, latent variables, emergent variables, or a mixture of all three are conceivable (Grace & Bollen, 2008; Schuberth et al., 2020; van Riel et al., 2017). The variables chosen to form the emergent ones are derived from a researcher's theory about the theoretical concept, i.e., about which ingredients are assumed to compose the forged concept. The most elaborate explanation of how to specify emergent variables can be found in Schuberth (in press). Considering the reflective and causal-formative measurement model, their specification is well elaborated in the existing literature. Hence, we refer to the studies of Bollen (2011) and Diamantopoulos (2011) for details about the specification of these models. Finally, the emergent and latent variables need to be linked in the structural model according to the researcher's theory. In doing so, different structural models, e.g., recursive and non-recursive ones, and other relationships, e.g., linear and non-linear ones, can be specified.

3.2 Model Identification

Once the model is specified, the next step must ensure that the specified model is identified. We mean by "a model is identified", that in theory, it is possible to obtain a unique set of model parameters from the variance-covariance matrix of the observable variables. Model identification is a necessary condition for trustworthy interpretation of the model parameter estimates (Marcoulides & Chin, 2013; Martin & Quintana, 2002).

Considering the composite model, identification can be achieved if the following two conditions are met: 1) the scale of the emergent variable has to be determined, 2) the emergent variable is not allowed to be isolated in the structural model, i.e., the emergent variable needs to have at least one antecedent or consequence, besides the variables forming the emergent variable. Determining the scale of an emergent variable can be achieved in various ways. Depending on the estimator employed, the variance can be, for instance, fixed directly, e.g., if ML is used (see Schuberth, in press), indirectly by scaling the weights to create an emergent variable with a unit variance. In both cases, the sign of the weights is not determined. Therefore, we recommend selecting a dominant indicator, i.e., the sign of the weights be chosen in such a way that the correlation between the dominant indicator and the emergent variable is positive (Henseler et al., 2016). Considering reflective and causal-formative measurement models, their identification is well covered in the literature such as Bollen (1989, Chapter 7) and Diamantopoulos (2011), and thus not elaborated here.

Finally, the structural model needs to be identified. Following Bollen (1989), the recursive structural models with uncorrelated error terms are always identified. For more complex structural models, e.g., non-recursive models, identification needs to be additionally assessed (e.g., Rigdon, 1995).

3.3 Model Estimation

A defining characteristic of composite-based SEM is the estimator employed. Composite-based SEM comprises two types of estimators: (i) estimators that employ composites as auxiliary variables to obtain the model parameter estimates, and (ii) estimators capable of estimating composite models. Considering the first category, PLSc and factor score regression with a correction for attenuation (Devlieger & Rosseel, 2017) are covered. Estimators such as PLS-PM, GSCA, and the ML estimator for composite models fall in the second category.

In choosing an estimator, a researcher should select an estimator that conforms to the model to be estimated, i.e., it produces unbiased and/or consistent estimates. On the one hand, if the model comprises only latent variables, estimators such as PLSc, GSCAm, and factor scores regression with a correction for attenuation provide estimates. On the other hand, if the model consists only of emergent variables, PLS-PM, GSCA, and the ML estimator for composite models are suitable choices. Since researchers usually study not only forged concepts, but also behavioral concepts, their research models most likely consist of both latent and emergent variables. In this case, estimators such as PLSc, IGSCA, or the ML estimator can be employed. Finally, the estimator must satisfy the characteristics of the structural model. For example, not all estimators allow for the estimation of feedback loops among constructs.

3.4 Model Assessment

Once the model has been estimated, it needs to be assessed. As is standard in SEM, the model assessment involves assessing the overall model fit and assessing the operationalized theoretical concepts, i.e., reflective measurement, causal-formative measurement, and composite model, including the relationships among the constructs.

Overall model fit assessment, i.e., investigating how well the model explains the data, is an integral part of SEM (Kline, 2015, p. 120). “If a model is consistent with reality, then the data should be consistent with the model” (Bollen, 1989, p. 68). The discrepancy between the empirical variance-covariance and the estimated model-implied variance-covariance matrix is considered in assessing the overall model fit. To do this, various ways have been proposed, which can be broadly organized into two categories: (i) statistical tests, and (ii) fit indices.

Various statistical tests have been proposed to assess the exact overall model fit (Yuan & Bentler, 1997, 1999). We need to emphasize that the suitability of a test depends on the employed estimator's statistical properties. If the normality assumption does not hold, i.e., the unobservable variables cannot be assumed to follow a multivariate normal distribution, robust versions of the chi-square test can be employed (for an overview, the interested reader is referred to Savalei, 2014). Alternatively, a non-parametric bootstrap-based test can be used to assess the exact overall model fit (Beran & Srivastava, 1985), as has also been proposed in the context of PLSc (Dijkstra & Henseler, 2015a).

Besides statistical tests, various fit indices have been proposed to assess the approximate fit of a model (e.g., Bentler & Bonett, 1980; Jöreskog & Sörbom, 1982). Although most of the fit indices were initially developed for latent variable models, they have recently been proposed for evaluating models containing emergent variables (Cho et al., 2020; Schuberth et al., Accepted). Potential candidates are the standardized root mean squared residual (SRMR, Bentler, 2006; Henseler et al., 2014), the root mean square error of approximation (RMSEA, Browne & Cudeck, 1993), the comparative fit index (CFI, Bentler, 1990), and the Tucker-Lewis index (TLI, Tucker & Lewis, 1973). In contrast to statistical tests, they are descriptive and typically compared to threshold values derived from simulation studies to judge the approximate fit of a model. However, we must point out that the common practice of comparing fit indices to thresholds has been criticized in the SEM literature (e.g., Marsh et al., 2004).

In assessing the model fit, a two-step assessment procedure has proven to be more advantageous than testing the complete model all at once (e.g., Anderson & Gerbing, 1988; Henseler et al., 2016). In the first step, the structural model's restrictions are ignored, and all constructs are allowed to be freely correlated. i.e., the structural model is saturated in this step. Subsequently, this model's fit is assessed to detect misspecifications in the reflective measurement, causal-formative measurement, and/or composite models. If the model contains latent variables only, a CFA is conducted. In contrast, if the model comprises only emergent variables, a CCA is conducted. If the model contains both latent and emergent variables, a confirmatory composite factor analysis (CCFA) is applied. In the second step, once the model's fit has been found acceptable in the first step, the initially specified model is assessed.

Once the overall model fit is acceptable, researchers should adopt the appropriate assessment criteria based on their operationalization of the theoretical concepts in their respective structural models. Importantly, the assessment criteria differ for reflective measurement models, causal-formative measurement models, and composite models. For an elaboration of the assessment criteria for the different models, we refer to Henseler et al. (2016), Benitez et al. (2020), and Henseler and Schuberth (2020b).

4 Illustrative Example

To demonstrate the relevance and usefulness of our proposed approach of determining a theoretical concept's operationalization by its nature, we show how it can be applied to a concrete IS research project. Instead of presenting an entirely new study, we review and re-analyze a well-received and influential[1] study in IS research, namely Rai et al. (2006). The study introduced *IT Infrastructure Integration for Supply Chain Management (SCM)* to the IS body of knowledge "as the degree to which a focal firm has established IT capabilities for the consistent and high-velocity transfer of supply chain-related information within and across its boundaries" (Rai et al., 2006, p. 229). This means that they conceptualize *IT Infrastructure Integration for SCM* as a particular kind of capability, namely an IT integration capability. Management scholars commonly regard capabilities as bundles of routines (Peng et al., 2008), which essentially means that they conceive of capabilities as forged concepts. Since capabilities "emerge" (Ellram et al., 2013, p. 31), are typically "built" or "acquired" (Barney, 2012, p. 4), and are understood to be developed, created, or produced (see, e.g., Andreu & Ciborra, 1996), they do not exist in nature. In particular, since *IT Infrastructure Integration for SCM* is IT-enabled, it can be defined as a forged concept. Notably, the charm of capabilities is that they are easily interpretable as attributes, in the sense of being more or less capable. This makes them eminently suitable as building blocks in theory building.

While Rai et al. (2006) showed how *IT Infrastructure Integration for SCM* can be created and that it has positive performance outcomes, they did not assess whether these performance

[1] As indicated by the absolute number of citations, as well as their distribution over time.

outcomes were attributable to the theoretical concept itself or merely to its constituents.[2] In other words, it remains unclear whether *IT Infrastructure Integration for SCM* acts as one whole, i.e., whether it makes sense to synthesize the forged concept. In addition, their model contains two more forged concepts, *Supply Chain Process Integration* and *Firm Performance*. In the following section, we will employ the four steps (i.e., model specification, model identification, model estimation, and model assessment) to assess whether empirical evidence speaks against the emergence of *IT Infrastructure Integration for SCM*.

4.1 Model Specification

Since *IT Infrastructure Integration for SCM*, *Supply Chain Process Integration*, and *Firm Performance* are understood to be forged concepts, we modeled them as emergent variables. According to Rai et al. (2006), *IT Infrastructure Integration for SCM* is made up of *Data Consistency* and *Cross-Functional Application Integration*; *Supply Chain Process Integration* is formed by *Financial Flow Integration*, *Physical Flow Integration*, and *Information Flow Integration*; and *Firm Performance* consists of *Operational Excellence*, *Revenue Growth*, and *Customer Relationships*.

In line with Rai et al. (2006), we hypothesized a direct effect of *IT Infrastructure Integration for SCM* on *Supply Chain Process Integration*, which in turn is hypothesized to affect *Firm Performance*. Moreover, two control variables were included as single-item constructs: *Consumer Demand Predictability* and *Firm Size*. The resulting model is depicted in Figure 1.

[2]In Rai et al.'s (2006) study, *IT Infrastructure Integration for SCM* is described as being operationalized through a causal-formative measurement model. However, both the construct definition and the actual modeling approach (partial least squares path modeling and principal component analysis) are in line with employing composites to model the theoretical concepts (Rigdon, 2016).

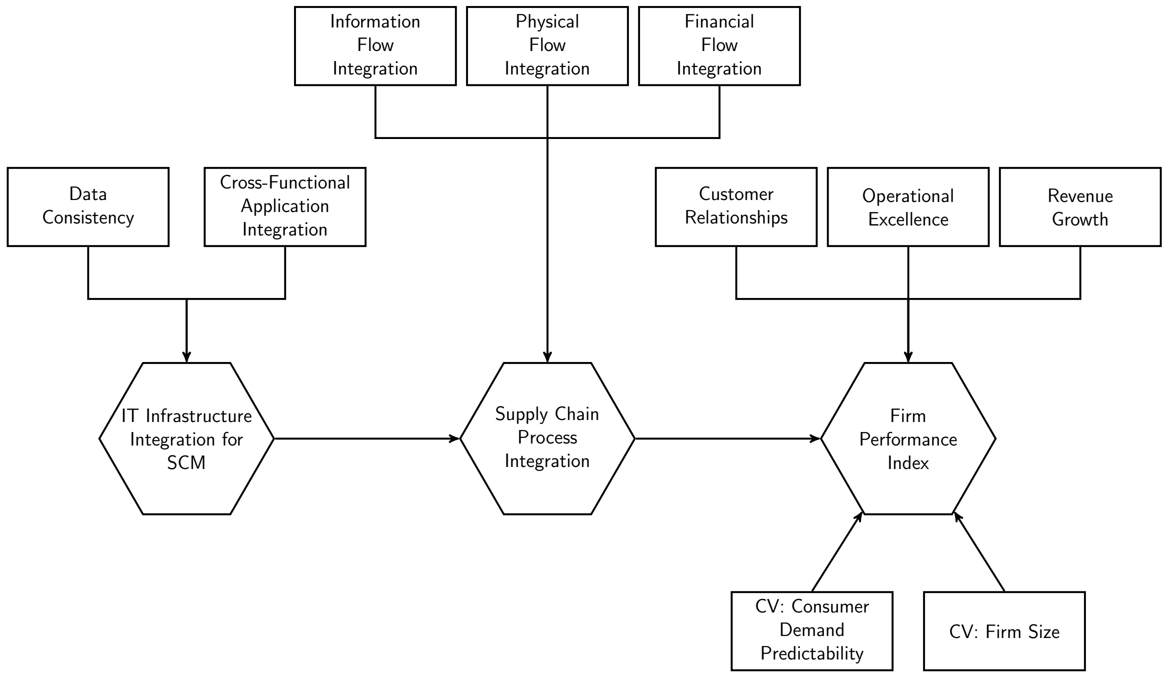


Figure 1: A model of *IT Infrastructure Integration for SCM*

4.2 Model Identification

We ensured that the specified model is identified, i.e., that the parameters can be theoretically, uniquely retrieved from the variance-covariance matrix of the observable variables. To do so, we relied on the identification rules formulated in Henseler (2021) and Schuberth (in press).

4.3 Model Estimation

As empirical data, we employed Rai et al.'s (2006) descriptive statistics and correlations from their study based on 110 observations collected in industrial organizations. The authors undertook multiple actions to ensure high data quality, such as screening the respondents' expertise and checking for non-response bias.

To obtain the model parameter estimates, we employed the full information maximum likelihood estimator as implemented in Mplus version 7.3 (Muthén & Muthén, 1998-2017).[3] The model estimation terminated normally.

The most important results of the model estimation are the model parameters, in our case the indicator weights and path coefficients. As shown in Table 2, virtually all estimates obtained from our estimation strongly resemble Rai et al.'s (2006) original estimates.

[3]The Mplus code is available from the corresponding author on request.

Table 2: Results of model estimation

Result	Rai et al. (2006) (PLS-PM)	Our analysis (ML)
Weights		
Data Consistency	.81	.81
Cross-Functional Application Integration	.28	.30
Information Flow Integration	.78	.77
Physical Flow Integration	.31	.28
Financial Flow Integration	.12	.17
Operational Excellence	.61	.62
Customer Relationships	.14	.13
Revenue Growth	.55	.55
Path to Supply Chain Process Integration		
IT Infrastructure Integration for SCM	.58	.58
Paths to Firm Performance		
Supply Chain Process Integration	.44	.43
Consumer Demand Predictability	−.04	−.04
Firm Size	−.09	.09
R^2		
Supply Chain Process Integration	.336	.333
Firm Performance	.186	.195

4.4 Model Assessment

Finally, we inspected the overall fit of the model, i.e., we empirically assessed the model's validity. The central model test result is a χ^2 value of 33.363, which corresponds to a p-value of .2226 (28 degrees of freedom). This implies that the model is not significantly inconsistent with the data, and the model should thus not be rejected. The commonly considered model fit indices gave the same picture: An SRMR of .071, an RMSEA of 0.042, a CFI of .968, and a TLI of .950 all indicate a well-fitting model.

Thus, overall, there is no empirical evidence for model misfit whatsoever. Consequently, we have no evidence against the emergent variables, which particularly holds for *IT Infrastructure Integration for SCM*. In combination with the positive direct and indirect effects of *IT Infrastructure Integration for SCM* on the variables of its nomological net, we can therefore conclude that it is useful to construct the forged concept *IT Infrastructure Integration for SCM*.

5 Discussion

Operationalizing theoretical concepts is a crucial task for IS researchers applying SEM. Initially, the reflective measurement model was proposed and became the dominant way of operationalizing. Recognizing that not every observable variable is necessarily a manifestation of a studied theoretical concept, i.e., a reflective measure, Petter et al. (2007) proposed the notion of the formative construct to the IS discipline. A formative construct is a latent variable affected by a set of observable variables as in the causal-formative measurement (Diamantopoulos, 2011). Besides the reflective and causal-formative measurement model, the study at hand elaborates a recently proposed third way of operationalizing theoretical concepts, namely, the composite model, which draws on emergent variables to represent theoretical concepts in the structural model (Benitez et al., 2020; Schuberth et al., 2018). Table 3 juxtaposes the three ways of operationalizing theoretical concepts.

Table 3: Characteristics of the composite, reflective measurement, and causal-formative measurement model

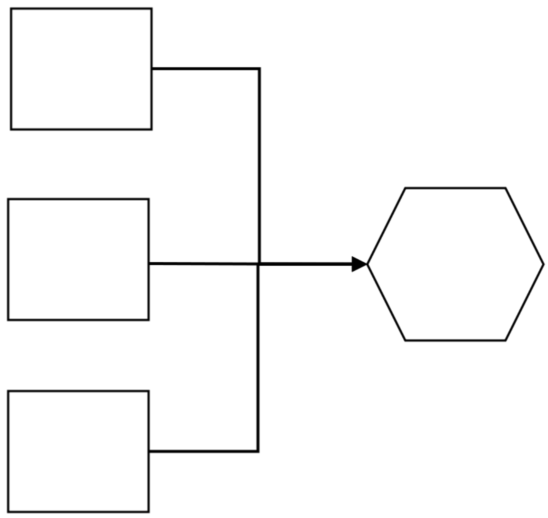
Characteristic	Composite model	Reflective measurement model	Causal-formative measurement model
Suitable to model type of theoretical concept	Forged concept	Behavioral concept	Behavioral concept
Type of construct	Emergent variable	Latent variable	Latent variable
Observable variables' role	Ingredients/elements	Consequences	Causes
Observable variables correlations	High correlations are possible	High correlations are expected	No reason to expect high correlations
Informative about observable variable's measurement error	Not informative	Informative	Not informative
Consequence of dropping an observable variable	Alters the construct's meaning	Does not alter the construct's meaning	Increases the error on construct level

In this study, we argue that the type of theoretical concept determines how a theoretical concept should be represented in the structural model. Theoretical concepts that are designed and assumed to emerge within their environment, so-called forged concepts, should be operationalized by the composite model comprising an emergent variable. In doing so, the observable variables that are assumed to compose the forged concept serve as ingredients, i.e., they have an assumed definitorial relationship with their concept, and removing an ingredient could potentially alter the construct's meaning. Although high correlations between the ingredients are common, they are not necessary. In the composite model we presented, the ingredients are assumed to be free from random measurement errors. To relax this assumption, researchers can manually adjust the reliability of their indicators (e.g., Mosier, 1943).

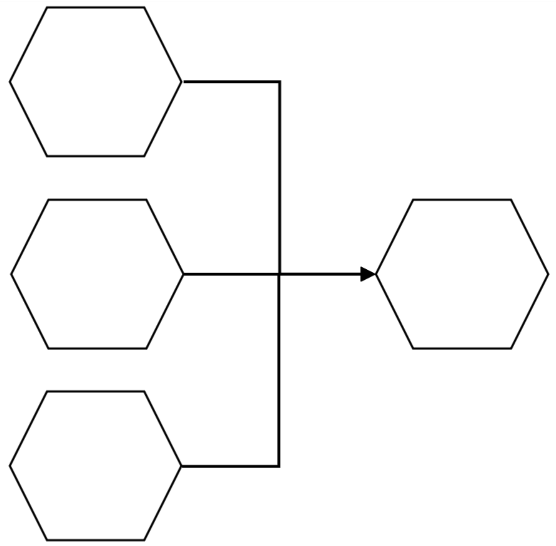
In contrast to forged concepts, behavioral concepts, assumed to exist in nature and typically measured by a set of observable variables, are best specified as latent variables by means of a reflective or causal-formative measurement model. In the reflective measurement model, the observable variables, i.e., reflective indicators, are assumed to be consequences of the theoretical concept; hence, removing an effect indicator should, in theory, not alter the meaning of the construct. Also, high correlations are expected between the indicators because they share the same underlying cause. In the causal-formative measurement model, the observable variables, i.e., causal indicators, are assumed to cause the theoretical concept; hence, removing a causal indicator will increase the error on the construct level even though it does not alter the construct's meaning (Aguirre-Urreta et al., 2016). Also, no reason to expect high correlations between the indicators.

To demonstrate the four steps of composite-based SEM and its use in modeling and assessing theories comprising forged concepts, we used Rai et al.'s (2006) empirical study as an illustrative case from the IS discipline. We examined IT Infrastructure Integration for SCM as a forged concept that is assumed to emerge from six attributes, namely as Information Flow Integration, Physical Flow Integration, Financial Flow Integration, Operational Excellence, Customer Relationships, and Revenue Growth. The results show no empirical evidence against an emergent variable. Moreover, we emphasize that our demonstrative case solely illustrates the steps of composite-based SEM. Therefore, several crucial steps involved in an empirical study have not been addressed, such as proper theory development, data collection, and assessment of the employed statistical tests' power. We simply built on the work of the original researchers.

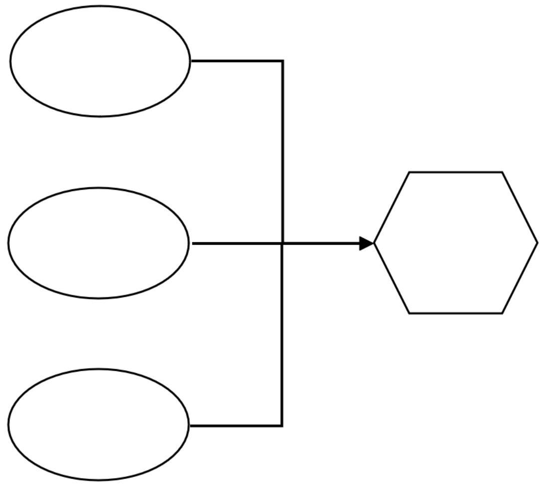
Overall, composite-based SEM opens new avenues for design-oriented IS researchers theorizing about forged concepts. For behavioral concepts, several guidelines have been proposed to assess their validity, including convergent validity, discriminant validity, and measurement invariance. However, for forged concepts such validity theories are less developed. Hence, it is up to future research to propose such theories and develop more sophisticated guidelines to assess forged concepts. Additionally, as shown in Figure 2, although we assume that emergent variables are made up of observable variables, in general, they can also be built by latent variables, emergent variables, or even a mixture of the two, i.e., a second-order emergent variable made of first-order latent variables, a second-order emergent variables made of first-order emergent variables, a second-order emergent variable made up of different types of variables (Schuberth et al., 2020; van Riel et al., 2017). Similarly, the complexity of models containing latent variables is currently much larger than those containing emergent variables; thus, it allows for the study of behavioral concepts in a broad range of contexts such as latent growth curve modeling (e.g., Muthén & Curran, 1997) and multilevel modeling (e.g., Rabe-Hesketh et al., 2004). We call for future research to examine whether these ideas can be transferred to forged concepts and how the proposed frameworks can be adopted to emergent variables.



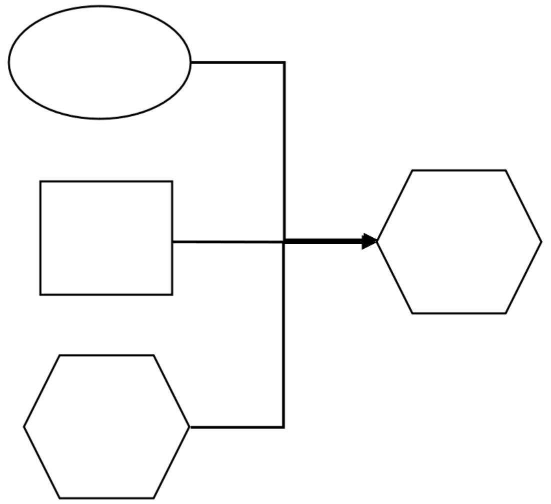
(a) An emergent variable made up of observable variables (default)



(b) An emergent variable made of emergent variables



(c) An emergent variable made of latent variables



(d) An emergent variable made of different types of variables

Figure 2: An emergent variable made up of different components

6 Conclusion

IS researchers study not only social units and entities but also technical ones. While latent variables were used to model the properties of the former, we propose forged concepts to be used to model the properties of the latter. We discuss how to operationalize forged concepts as emergent variables in the composite model, distinct from the causal-formative and reflective measurement models. Most importantly, we offer CCA or composite-based SEM as the statistical method to assess the composite models. To help IS researchers assess composite models, we provided an empirical example from the IS literature and detailed four steps to follow: model specification, model identification, model estimation, and model assessment. We hope this new conceptualization and operationalization will trigger researchers to consider forged concepts as another type of theoretical concepts and CCA or composite-based SEM as an expansion to their toolset to achieve a better understanding of the theoretical concept under study and ultimately contribute to research and practice.

References

Aguirre-Urreta, M. I., Rönkkö, M., & Marakas, G. M. (2016). Omission of causal indicators: Consequences and implications for measurement. *Measurement: Interdisciplinary Research and Perspectives*, *14*(3), 75–97. https://doi.org/10.1080/15366367.2016.1205935

Ahuja, M., Chudoba, K. M., Kacmar, C. J., Mcknight, D. H., & George, J. F. (2007). It road warriors: Balancing work-family conflict, job autonomy, and work overload to mitigate turnover intentions. *MIS Quarterly*, *31*(1), 1–17. https://doi.org/10.2307/25148778

Allen, M. W., Armstrong, D. J., Reid, M. F., & Riemenschneider, C. K. (2009). IT employee retention: Employee expectations and workplace environments. *SIGMIS CPR '09: Proceedings of the special interest group on management information system's* 47^{th} *annual conference on Computer personnel research*, 95–100.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Andreu, R., & Ciborra, C. (1996). Organisational learning and core capabilities development: The role of it. *The Journal of Strategic Information Systems*, *5*(2), 111–127. https://doi.org/10.1016/s0963-8687(96)80039-4

Armstrong, D. J., Brooks, N. G., & Riemenschneider, C. K. (2015). Exhaustion from information system career experience: Implications for turn-away intention. *MIS Quarterly*, *39*(3), 713–727. https://doi.org/10.25300/misq/2015/39.3.10

Armstrong, D. J., Riemenschneider, C., Buche, M., & Armstrong, K. (2018). Mitigating turnover intentions: Are all it workers warriors? *AIS Transactions on Replication Research*, *4*, 1–20. https://doi.org/10.17705/1atrr.00030

Bagozzi, R. P., & Phillips, L. W. (1982). Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly*, *27*(3), 459–489. https://doi.org/10.2307/2392322

Barney, J. B. (2012). Purchasing, supply chain management and sustained competitive advantage: The relevance of resource-based theory. *Journal of Supply Chain Management*, *48*(2), 3–6. https://doi.org/10.1111/j.1745-493x.2012.03265.x

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*, *57*(2), 103168. https://doi.org/10.1016/j.im.2019.05.003

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Multivariate Software, Inc.

Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, *13*(1), 95–115. https://doi.org/10.1214/aos/1176346579

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. University of North Carolina Press.

Bollen, K. A. (1984). Multiple indicators: Internal consistency or no necessary relationship? *Quality and Quantity*, *18*(4), 377–385. https://doi.org/10.1007/bf00227593

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35*(2), 359–372.

Bollen, K. A., & Diamantopoulos, A. (2017a). In defense of causal–formative indicators: A minority report. *Psychological Methods*, *22*(3), 581–596.

Bollen, K. A., & Diamantopoulos, A. (2017b). Notes on measurement theory for causal-formative indicators: A reply to Hardin. *Psychological Methods*, *22*(3), 605–608.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspective*, *6*(1-2), 25–53. https://doi.org/10.1080/15366360802035497

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219.

Brown, & Venkatesh. (2005). Model of adoption of technology in households: A baseline model test and extension incorporating household life cycle. *MIS Quarterly*, *29*(3), 399–426. https://doi.org/10.2307/25148690

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). Sage.

Cho, G., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika*, *47*, 243–272. https://doi.org/10.1007/s41237-019-00098-0

Cho, G., Hwang, H., Sarstedt, M., & M., R. C. (2020). Cutoff criteria for overall model fit indexes in generalized structured component analysis. *Journal of Marketing Analytics*, *8*, 189–202.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis. *Methodology*, *13*, 31–38.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335–358.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277.

Dijkstra, T. K. (2013). Composites as factors: Canonical variables revisited. *Working Paper*. https://doi.org/10.13140/RG.2.1.3426.5449

Dijkstra, T. K. (2017). A perfect match between a model and a mode. In H. Latan & R. Noonan (Eds.), *Partial least squares path modeling: Basic concepts, methodological issues and applications* (pp. 55–80). Springer.

Dijkstra, T. K., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, *81*, 10–23. https://doi.org/10.1016/j.csda.2014.07.008

Dijkstra, T. K., & Henseler, J. (2015b). Consistent partial least squares path modeling. *MIS Quarterly*, *39*(2), 297–316.

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On product uncertainty in online markets: Theory and evidence. *MIS Quarterly*, *36*(2), 395–426.

Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society*, *51*(3), 599–635.

Ellram, L. M., Tate, W. L., & Feitzinger, E. G. (2013). Factor-market rivalry and competition for supply chain resources. *Journal of Supply Chain Management*, *49*(1), 29–46. https://doi.org/10.1111/jscm.12001

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

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51–90. https://doi.org/10.2307/30036519

Grace, J. B., & Bollen, K. A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistics*, *15*(2), 191–213. https://doi.org/10.1007/s10651-007-0047-7

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science*, *45*(5), 616–632. https://doi.org/10.1007/s11747-017-0517-x

Hardin, A. (2017). A call for theory to support the use of causal-formative indicators: A commentary on Bollen and Diamantopoulos (2017). *Psychological Methods*, *22*(3). https://doi.org/10.1037/met0000115

Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analyses. *Sociological Methods & Research*, *1*(2), 147–173. https://doi.org/10.1177/004912417200100201

Henseler, J. (2015a). Is the whole more than the sum of its parts.

Henseler, J. (2015b). *Is the whole more than the sum of its parts? On the interplay of marketing and design research*. Universiteit Twente.

Henseler, J. (2017). Bridging design and behavioral research with variance-based structural equation modeling. *Journal of Advertising*, *46*(1), 178–192. https://doi.org/10.1080/00913367.2017.1281780

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

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Hair, J. F., Hult, G. T. M., Calantone, R. J., & et al. (2014). Common beliefs and reality about PLS. *Organizational Research Methods*, *17*(2), 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*(1), 2–20. https://doi.org/10.1108/imds-09-2015-0382

Henseler, J., & Schuberth, F. (2020a). Auxiliary theories. *Composite-based structural equation modeling: Analyzing latent and emergent variables* (pp. 25–37). Guilford Press.

Henseler, J., & Schuberth, F. (2020b). Using confirmatory composite analysis to assess emergent variables in business research. *Journal of Business Research*, *120*, 147–156. https://doi.org/10.1016/j.jbusres.2020.07.026

Ho, V. T., Ang, S., & Straub, D. (2003). When subordinates become it contractors: Persistent managerial expectations in it outsourcing. *Information Systems Research*, *14*(1), 66–86. https://doi.org/10.1287/isre.14.1.66.14764

Hubona, G. S., Schuberth, F., & Henseler, J. (2021). A clarification of confirmatory composite analysis (cca). *International Journal of Information Management*, *61*, 102399. https://doi.org/10.1016/j.ijinfomgt.2021.102399

Hwang, H., Cho, G., Jung, K., Falk, C., Flake, J., Jin, M. J., & Lee, S. H. (2021). An approach to structural equation modeling with both factors and components: Integrated generalized structured component analysis. *Psychological Methods*, *26*(3), 273–294.

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81–99. https://doi.org/10.1007/bf02295841

Jarvis, C. B., Mackenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218. https://doi.org/10.1086/376806

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: An empirical study. *MIS Quarterly*, *34*(3), 549–566.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, *57*(2), 239–251.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*(351), 631–639.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, *19*(4), 404–416.

Joseph, D., Ng, K.-Y., Koh, C., & Ang, S. (2007). Turnover of information technology professionals: A narrative review, meta-analytic structural equation modeling, and model development. *MIS Quarterly*, *31*(3), 547–577. https://doi.org/10.2307/25148807

Karimi, J., Somers, T. M., & Gupta, Y. P. (2004). Impact of environmental uncertainty and task characteristics on user satisfaction with data. *Information Systems Research*, *15*(2), 175–193. https://doi.org/10.1287/isre.1040.0022

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Lance, C. E., Cornwell, J. M., & Mulaik, S. A. (1988). Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research*, *23*(2), 171–187. https://doi.org/10.1207/s15327906mbr2302_3

Law, K. S., & Wong, C.-S. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management*, *25*(2), 143–160. https://doi.org/10.1016/s0149-2063(99)80007-5

Lewis, W., Agarwal, R., & Sambamurthy, V. (2003). Sources of influence on beliefs about information technology use: An empirical study of knowledge workers. *MIS Quarterly*, *27*(4), 657–678. https://doi.org/10.2307/30036552

Marcoulides, G. A., & Chin, W. W. (2013). You write, but others read: Common methodological misunderstandings in PLS and related methods. In H. Abdi, V. E. Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New perspectives in partial least squares and related methods* (pp. 31–64). Springer.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Martin, E. S., & Quintana, F. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, *16*, 99–106.

Moquin, R., Riemenschneider, C. K., & Wakefield, R. L. (2019). Psychological contract and turnover intention in the information technology profession. *Information Systems Management*, *36*(2), 111–125. https://doi.org/10.1080/10580530.2019.1587574

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, *8*(3), 161–168. https://doi.org/10.1007/bf02288700

Müller, T., Schuberth, F., & Henseler, J. (2018). PLS path modeling - a confirmatory approach to study tourism technology and tourist behavior. *Journal of Hospitality and Tourism Technology*, *9*(3), 249–266.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*(4), 371–402.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus* (8th). Muthén & Muthén.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Peng, D. X., Schroeder, R. G., & Shah, R. (2008). Linking routines to operations capabilities: A new perspective. *Journal of Operations Management*, *26*(6), 730–748. https://doi.org/10.1016/j.jom.2007.11.001

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623–656. https://doi.org/10.2307/25148814

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167–190.

Rademaker, M. E., Schuberth, F., & Dijkstra, T. K. (2019). Measurement error correlation within blocks of indicators in consistent partial least squares. *Internet Research*, *29*(3), 448–463. https://doi.org/10.1108/intr-12-2017-0525

Rai, Patnayakuni, & Seth. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS Quarterly*, *30*(2), 225–246. https://doi.org/10.2307/25148729

Rasoolimanesh, S. M., Noor, S. M., Schuberth, F., & Jaafar, M. (2019). Investigating the effects of tourist engagement on satisfaction and loyalty. *The Service Industries Journal*, *39*(7-8), 559–574. https://doi.org/10.1080/02642069.2019.1570152

Reichenbach, H. (1956). *The direction of time*. University of California Press.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, *45*(5-6), 341–358. https://doi.org/10.1016/j.lrp.2012.09.010

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, *30*(3), 359–383.

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

Rutner, P. S., Hardgrave, B. C., & McKnight, D. H. (2008). Emotional dissonance and the information technology professional. *MIS Quarterly*, *32*(3), 635–652.

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. https://doi.org/10.1016/j.jbusres.2016.06.007

Savalei, V. (2014). Understanding robust corrections in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(1), 149–160. https://doi.org/10.1080/10705511.2013.824793

Schuberth, F. (2021). Confirmatory composite analysis using partial least squares: Setting the record straight. *Review of Managerial Science*, *15*, 1311–1345.

Schuberth, F. (in press). The henseler-ogasawara specification of composites in structural equation modeling: A tutorial. *Psychological Methods*. https://doi.org/10.1037/met0000432

Schuberth, F., Henseler, J., & Dijkstra, T. K. (2018). Confirmatory composite analysis. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02541

Schuberth, F., Rademaker, M. E., & Henseler, J. (Accepted). Assessing the overall fit of composite models estimated by partial least squares path modeling. *European Journal of Marketing*.

Schuberth, F., Rademaker, M. E., & Henseler, J. (2020). Estimating and assessing second-order constructs using PLS-PM: The case of composites of composites. *Industrial Management & Data Systems*, *120*(12), 2211–2241. https://doi.org/10.1108/imds-12-2019-0642

Schumacker, R. E., & Lomax, R. G. (2009). *A beginner's guide to structural equation modeling* (3rd). Routledge.

Thatcher, J. B., Wright, R. T., Sun, H., Zagenczyk, T. J., & Klein, R. (2018). Mindfulness in information technology use: Definitions, distinctions, and a new measure. *MIS Quarterly*, *42*(3), 831–848.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, *28*(4), 695–704. https://doi.org/10.2307/25148660

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*(3), 459–477. https://doi.org/10.1108/IMDS-07-2016-0286

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*(4), 490–495.

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16*(1), 85–102. https://doi.org/10.1287/isre.1050.0042

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. Blalock, A. Aganbegian, F. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology* (pp. 307–357). Academic Press.

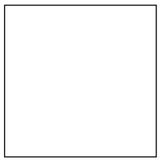
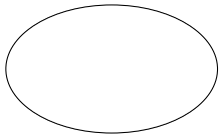
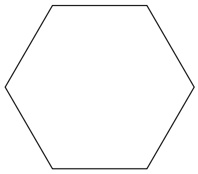
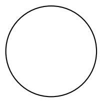
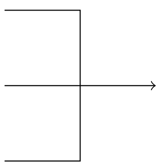
Xu, J., Benbasat, I., & Cenfetelli, R. T. (2013). Integrating service quality with system and information quality: An empirical test in the e-service context. *MIS Quarterly*, *37*(3), 777–794. https://doi.org/10.25300/misq/2013/37.3.05

Yuan, K.-H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, *92*(438), 767–774. https://doi.org/10.1080/01621459.1997.10474029

Yuan, K.-H., & Bentler, P. M. (1999). F tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics*, *24*(3), 225–243. https://doi.org/10.3102/10769986024003225

Zhang, X., & Venkatesh, V. (2017). A nomological network of knowledge management system use: Antecedents and consequences. *MIS Quarterly*, *41*(4), 1275–1306. https://doi.org/10.25300/misq/2017/41.4.12

A Elements of structural equation models

Symbol	Explanation
	Observable variable
	Latent variable
	Emergent variable
	Disturbance term
	Causal relationship
	Definitorial relationship