The Henseler-Ogasawara Specification of Composites in Structural Equation Modeling: A Tutorial

Florian Schuberth
Department of Design Production and Management, Faculty of Engineering Technology, University of Twente

Abstract

 Structural equation modeling (SEM) is a versatile statistical method that should theoretically be able to emulate all other methods that are based on the general linear model. In practice, however, researchers using SEM encounter problems incorporating composites into their models. In this tutorial article, I present a specification for SEM that was recently sketched by Henseler (2021) to incorporate composites in structural models. It draws from the same idea that was proposed in the context of canonical correlation analysis to express a set of observed variables forming a composite by a set of synthetic variables (Ogasawara, 2007), which were labeled by Henseler (2021) as emergent and excrescent variables. An emergent variable is a linear combination of variables that is related to other variables in the structural model, whereas an excrescent variable is a linear combination of variables that is unrelated to all other variables in the structural model. This approach is advantageous over existing approaches, as it allows drawing on all existing developments in SEM, such as testing parameter estimates, testing for overall model fit and dealing with missing values. To demonstrate the presented approach, I conduct a small scenario analysis. Moreover, SEM applying the presented specification is used to reestimate an empirical example from Hwang et al. (2021). Finally, this article discusses avenues for future research opened by this approach for SEM to study composites.

Translational abstract

Structural equation modeling (SEM) is a statistical method that allows researchers great flexibility in specifying their models. Yet, researchers face difficulties incorporating composite into their models. To address this issue, in this tutorial paper I present the Henseler-Ogasawara (H-O) specification for composites, which gives researchers the same flexibility that they are accustomed from modeling with latent variables. The H-O specification draws from the idea to express a set of observed variables forming a composite by a set of emergent and excrescent variables. An emergent variable is a linear combination of variables that is related to other variables in the structural model, whereas an excrescent variable is a linear combination of variables that is unrelated to all other variables in the structural model. This approach is advantageous over existing approaches, as it allows drawing on all existing developments in SEM, such as testing parameter estimates, testing for overall model fit and dealing with missing values.

Keywords: components, composites, emergent variables, excrescent variables, structural equation modeling

Supplemental materials: https://doi.org/10.1037/met0000432.supp

Structural equation modeling (SEM) with latent variables is a widely acknowledged and highly appreciated approach among researchers from social sciences (e.g., Bollen, 1989; Kline, 2015). It emerged conceptually from factor analysis (Pearson, 1901) and path analysis (Wright, 1918); see Tarka, 2018 for information about the development of SEM. In its current form, it combines the benefits of various types of multivariate data analysis and shows a high degree of modeling flexibility (Graham, 2008; Hoyle, 2012). Overall, SEM is a confirmatory approach that aims at modeling and testing theories comprising abstract concepts (Bagiotti & Phillips, 1982; Edwards & Bagiotti, 2000). These abstract concepts are typically represented by common factors, that is, latent variables, which are connected via a structural model. Finally, overall model fit assessment in the form of statistical tests and fit indices can be employed to obtain empirical evidence against a researcher’s theory (Schermelleh-Engel et al., 2003).

Over the last decades, composites—that is, linear combinations of variables—have gained the attraction of researchers using SEM. Their use can be roughly divided into two categories. First, composites, for example, in form of factor scores, are often used...
as approximations for common factors. However, it is well known that this approach produces inconsistent estimates (Bollen, 1989; Cohen et al., 1990). To overcome this drawback, various approaches have been developed that correct for the reliability of the composite scores to obtain consistent estimates for models relating latent variables (e.g., Devlieger & Rosseel, 2017; Dijkstra & Schermelleh-Engel, 2014; Wall & Ameniya, 2003). Second, composites are regarded as a viable means of modeling abstract concepts that are formed or defined and not measured by their variables (e.g., Cohen et al., 1990; Edwards, 2001; Edwards & Bagozzi, 2000; Henseler, 2017, 2021; Henseler & Schuberth, 2020; McDonald, 1996). In this case, the common factor model which assumes that the observed variables are random-measurement-error-prone manifestations of the underlying concept is of little value and the composite appears to be a more natural way of representing these concepts in the statistical model. Examples of such concepts are general ability (Conway & Kovacs, 2015; Eid et al., 2017) and marketing mix (Fornell & Bookstein, 1982) which have been proposed to be modeled as constructs that are composed, that is, composites. Similarly, stress and plant community structure can be viewed as concepts that emerge as a linear combination of variables (Antoninka et al., 2011; Hancock, 1997). Further examples are given by socioeconomic status which can be modeled as composite as that is defined in terms of occupation, education and income (Edwards & Bagozzi, 2000); job performance which should be conceptualized as a composite (Murphy & Shiarlly, 1997); and gene or brain regions that can be regarded as biological composites of single nucleotide polymorphisms or voxels, respectively (Jung et al., 2012, 2016; Romdhani et al., 2015). In this tutorial, I focus on composites as a means of representing concepts that are assumed to be formed by a set of variables.

Although early literature on SEM and common factor analysis (CFA) has already mentioned the possibility of SEM with composites (e.g., Schönemann & Steiger, 1976), the application of composites in SEM is still rather limited, and dedicated techniques emerged outside the realm of classical SEM to analyze composites embedded in a structural model. As a consequence, researchers have begun to distinguish between factor-based and composite-based SEM (e.g., Rigdon, 2012). While factor-based SEM refers to traditional SEM with latent variables, composite-based SEM refers to approaches that link composites in a structural model. The two arguably most widespread composite-based SEM approaches are partial least squares path modeling (PLS-PM; Wold, 1975), also known as partial least squares structural equation modeling (PLS-SEM; Hair et al., 2011) and generalized structured component analysis (GSCA; Hwang & Takane, 2004; Gläss, 1988). However, in contrast to traditional SEM, these techniques (still) show several limitations. For example, in PLS-PM, it is not possible to constrain parameters, which limits its modeling flexibility. Similarly, in GSCA, it is currently not possible to test the overall model fit, which limits its use for explanatory modeling and theory testing.

To incorporate composites in SEM, several specifications have been proposed such as a two-step approach, a one-step approach, and an approach based on a multiple indicators, multiple causes (MIMIC) model (Bagozzi et al., 1981). However, all these approaches entail several drawbacks, which limit their applicability of including composites in structural models and therefore the capability of SEM to study composites. For instance, the two-step approach creates composites outside the model and thus ignores the characteristics of the composite model, namely, that all the information between the variables forming the composite and other variable is fully conveyed by the composite (Dijkstra, 2017). Moreover, although the approach based on the MIMIC model was proposed to conduct a canonical correlation analysis within the SEM framework and is thus in principle capable of including two composites, in its current form, it misses a substantial amount of the flexibility known from traditional SEM, for example, it is not clear how to include more than two composites in the model.

Only recently, a new way of specifying composites in SEM was sketched by Henseler (2021), which overcomes the shortcomings of extant specifications to include composites in SEM. This specification largely resembles a model specification proposed by Ogasawara (2007; which was later extended by Gu et al., 2019) to conduct a canonical correlation analysis using SEM. It draws from the idea of modeling a set of observed variables forming a composite by a set of synthetic variables and expressing the relationship between the observed variables and the synthetic variables by loadings instead of weights. As a consequence, this model specification allows researchers to study and model composites in SEM with the same rigor and flexibility as they are accustomed from modeling with common factors.

The purpose of this tutorial article is to elaborate this recently proposed approach and to demonstrate its flexibility regarding modeling with composites in SEM. In doing so, it contributes to unifying composite-based and factor-based SEM and makes the developments of SEM, such as tests for the overall model fit, accessible to SEM with composites including the application of SEM software packages such as lavaan (Rosseel, 2012) and Mplus (Muthén & Muthén, 1998–2017). The remainder of the article is structured as follows. In the next section, I give an overview of existing approaches known from SEM to incorporate composites and highlight their shortcomings. Subsequently, I elaborate on the specification sketched by Henseler (2021) to model with composites in SEM. Furthermore, to demonstrate this innovative way of model specification and to show its advantage over the conventional way of specifying composites in SEM, the specification is applied to both artificial data and a real case. Finally, in the last section I conclude the article and give an outlook for future research.

### Extant Approaches to Incorporate Composites in Structural Models

In SEM, a composite $C$, which is also called component, is defined as a weighted sum (with weights $w$) of observed variables $x$ (see, e.g., Bollen & Bauldry, 2011) and can be written as follows:

$$ C = x'w $$

(1)

While this equation is characterized by an impressive simplicity, it poses a challenge to researchers using SEM who want to

---

1. The literature also labels this type of SEM as component-based SEM (e.g., Hwang et al., 2021). In the following, I will use composite-based nomenclature as an umbrella for both composite-based and component-based SEM.

2. In general, composites can also be formed by a set of common factors, composites, or a mixture of both (Edwards, 2001; Schuberth et al., 2020; Van Riel et al., 2017). However, in this article I limit the scope to composites formed by observed variables.
incorporate composites in their structural equation models. The reason for this difficulty can be found in the fact that there is a mismatch between the parameters involved in a structural equation model and those used to form a composite. This can be illustrated by considering the model-implied variance-covariance matrix of the observed variables for a linear structural equation model, which takes the following form (see, e.g., McDonald, 1996):

$$\Sigma(\theta) = \Lambda(I - B)^{-1}\Psi(I - B)^{-1}\Lambda' + \Theta_e$$  \hspace{1cm} (2)

In this equation, $\Sigma(\theta)$ denotes the model-implied variance-covariance matrix of the observed variables, which depends on the parameters $\theta$ that are captured in the following matrices: The matrix $A$ contains the loadings, the matrix $\Psi$ contains the variances and covariances of the structural error terms and the exogenous latent variables, the matrix $B$ contains the structural coefficients, and the matrix $\Theta_e$ contains the variances and covariances of the random measurement errors. Moreover, $I$ is a unit matrix of the same dimension as $B$. Obviously, the set of model parameters $\theta$ does not include any weights.

Considering the SEM literature, it offers three approaches regarding how researchers can incorporate composites into structural equation models: (a) a two-step approach, which separates the creation of composites from conducting SEM; (b) a one-step approach, which conducts SEM with reparameterized weights; and (c) an approach based on a MIMIC model.

The Two-Step Approach

The two-step approach entails that analysts create composites of observed variables before they start with SEM and subsequently use the composite scores as input for SEM. This approach is often employed if researchers make use of sum scores (i.e., linear combinations of observed variables using equal weights) or indices with other predefined weights. The two-step approach is also the method of choice if item parceling is applied (see, e.g., Little et al., 2002, 2013).

If composite scores are used as input data for SEM instead of modeling the composite directly in SEM (see the one-step approach in the following section), the amount of available moments to estimate and assess the model is reduced. As a consequence, the two-step approach can influence parameter estimates, affect the statistical power of the test for exact overall model fit, and provide opportunities for obfuscating model missfit. Moreover, following this approach, weights are not regarded as model parameters and are thus not estimated, but rather determined. Consequently, statistical inference about the weight estimates is not accessible to researchers. Similarly, composites remain unassessed with regard to overall model fit assessment.

The One-Step Approach

The problems with the two-step approach can be avoided if the creation of composites is included in the structural equation model, as is the case in the one-step approach. The one-step approach treats the weights as structural coefficients: $\eta = x' b$ (see e.g., Grace et al., 2010). One interpretation of this approach is that the observed variables $x$ forming the composite are elevated to single-indicator constructs that fully explain an endogenous latent variable $\eta$, that is, the variance of its disturbance term is fixed to zero. The one-step approach works well if the structural model contains only one composite in an exogenous position, that is, it is not affected by any variable in the structural model. However, in all other instances, the approach turns out to be problematic.

First, analysts face ambiguity in how to treat the covariances among the observed variables forming composites if a structural equation model contains more than one composite. While there seems to be agreement that the covariances among the observed variables of one composite should be free model parameters, there are opposing views on how to deal with the covariances between observed variables belonging to different composites. For example, MacCallum and Browne (1993) fix these covariances to zero, whereas Grace and Bollen (2008) specify them as free parameters. The difference between the two specifications can be substantial in terms of degrees of freedom, parameter estimates, and goodness of model fit.

Second, structural equation models with endogenous composites are not identified in the case of the one-step approach (MacCallum & Browne, 1993). This also explains why the two-step approach is often used, for instance, for item parceling; it is not possible to specify composites of observed variables using the one-step approach that at the same time are caused by variables other than the variables making up the composite.

Third, the one-step approach models the relationship between a composite and its observed variables analogous to a regular (causal) relationship; only the variance of the disturbance term is fixed to zero. However, this constraint needs to be inevitably relaxed if a composite is in an endogenous position in the structural model, as it is not reasonable to assume that the independent constructs together with the observed variables forming the composite fully cause the variation in the endogenous composite.

Fourth, the fact that the one-step approach treats composites automatically as endogenous constructs prevents researchers from modeling covariances between composites because only the covariances between the disturbance terms of the composites can be specified. Thus, models with composites that covary with other variables cannot be specified.

The MIMIC Approach

Bagoszi et al. (1981) showed that a type of MIMIC model, as depicted in Figure 1, can be used to conduct a canonical correlation analysis in SEM. Given that canonical correlation analysis is used to analyze pairs of correlated composites, this specification might provide an opportunity for SEM to model with composites in a flexible way.

In fact, the MIMIC approach overcomes the problem of the two-step approach that the model-implied variance-covariance matrix of the observed variables is not correctly determined. However, the MIMIC approach suffers from two other problems.

The first problem of the MIMIC approach is overparameterization. In many situations, the MIMIC approach employs more (free) parameters than actually necessary to specify a set of interrelated composites. The problem is illustrated by means of the model depicted in Figure 1. In the MIMIC approach, this model has 18 free parameters: three contained in $A$, three in $B$, six in $\Psi$, and six in $\Theta_e$. However, as Figure 2 shows, when specified outside of SEM, only 17 parameters are actually required to describe a model of two interrelated composites, each composed by three observed
This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The second problem involves the generalization of the MIMIC approach beyond canonical correlation analysis to allow for a flexible way of modeling with composites in SEM. The first composite made up of $x_1$ to $x_3$ is apparent in the MIMIC model and modeled via the one-step approach described above. However, the second composite—that is, the linear combination of $x_4$ to $x_6$—is not directly modeled but expressed by cross-loadings of $x_4$, $x_5$, and $x_6$ on the first composite. While the problems of the one-step approach have been previously elaborated, it is not clear, nor do Bagozzi et al. (1981) provide any suggestions on how expressing the composite via cross-loadings of its variables on other variables can be used to model composites in a flexible way. For instance, it is not clear how a composite can be modeled in an endogenous position of larger structural model using this specification. Moreover, this specification requires additional steps to retrieve the relevant information about the second composite. Most importantly, the covariance between the two composites—often the focal piece of information from a model of interrelated composites—does not directly appear as a parameter.

Another look at the MIMIC model reveals similarities to the “composite factor model” that was introduced recently to represent composites (Henseler et al., 2014). The composite factor model is a modified common factor model in which random measurement errors can freely covary. In principle, this specification may be used to model composites in an endogenous position of the structural model. However, this model shows a severe problem; namely, it is generally not identified even if the common factor is embedded in a structural model (McIntosh et al., 2014). Against this background, it is concluded that the MIMIC approach including the composite factor model does not allow for a flexible way of modeling with composites in SEM.

Approaches From Outside the Realm of Traditional SEM

The difficulties researchers face with modeling composites in traditional SEM seem to lie at an even deeper layer: The composite model has simply not been fully recognized as a statistical model with meaningful and testable implications. Traditionally, composites have merely been regarded as mathematical operations or tools for dimension reduction and not as statistical entities representing concepts. For instance, researchers do not ascribe any conceptual unity to the variables forming a composite (Bollen & Diamantopoulos, 2017), and methods dedicated to composites, such as canonical correlation analysis, do not have a history of assessing the goodness of model fit (Fan, 1997). In addition, yes, a single composite has no testable implications; its parameters are not even identified because, in principle, any weights can be used to form the composite. However, once a composite is related to other variables in a statistical model, the situation is quite different.

Considering SEM techniques that relate composites, which emerged outside the realm of traditional SEM, the composite model has already been recognized as a statistical model. Specifically, 

$$
\begin{bmatrix}
\phi_{11} \\
\phi_{12} \\
\phi_{13} \\
\phi_{21} \\
\phi_{22} \\
\phi_{23} \\
\phi_{31} \\
\phi_{33}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5 \\
x_6
\end{bmatrix}
\begin{bmatrix}
\beta_{41} \\
\beta_{42} \\
\beta_{43}
\end{bmatrix}
\begin{bmatrix}
\lambda_{41} \\
\lambda_{45} \\
\lambda_{46}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_3 \\
\varepsilon_4 \\
\varepsilon_5 \\
\varepsilon_6
\end{bmatrix}
$$

Figure 1

**Specification Proposed by Bagozzi et al. (1981) to Model Two Interrelated Composites**

variables: six variances of the observed variables, six covariances among the observed variables, four weights, and last but not least, the covariance between the composites. As a consequence of over-parameterization, the number of degrees of freedom in the MIMIC approach is typically too low, as the model is too generous, which contradicts the principle of Ockham’s razor.

$3$ Note that the weights are scaled to ensure that each composite has unit variance (alternatively, one weight per composite can be fixed to a value not equal to zero). As a consequence, one weight of each composite is not a free parameter but can be expressed as a function of the other two weights, covariances among the observed variables, and the observed variables’ variances. For instance, the third weight of the first composite can be obtained by solving the following equation for $w_{31}$:

$$
\begin{bmatrix}
w_{11} \\
w_{21} \\
w_{31}
\end{bmatrix}
\begin{bmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{bmatrix}
\begin{bmatrix}
w_{11} \\
w_{21} \\
w_{31}
\end{bmatrix}
= 1.
$$

$$
\begin{bmatrix}
w_{11} \\
w_{21} \\
w_{31}
\end{bmatrix}
\begin{bmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{bmatrix}
\begin{bmatrix}
w_{11} \\
w_{21} \\
w_{31}
\end{bmatrix}
= 1.
$$
composites embedded in a larger statistical model can have testable implications, and this creates an opportunity that has so far been largely overlooked in the traditional SEM literature (e.g., Cho & Choi, 2020; Dijkstra, 2013, 2017). Concretely, the composite model signifies that the observed variables forming a composite only act through their composite and not as single entities. This implies that the covariances between the observed variables forming a composite and other variables in the model always exhibit the same proportion, which is similar to the causal-formative measurement model (Bollen & Davis, 2009; Franke et al., 2008). In other words, the covariance matrix of observed variables belonging to a composite and another variable is of rank one.

The arguably most widespread techniques that relate composites in structural models are PLS-PM (Wold, 1975, 1982) and GSCA (Hwang & Takane, 2004, 2014). Originally, both were developed to deal with structural models solely containing composites. Only recently, PLS-PM and GSCA have been extended to deal with structural models containing both common factors and composites, namely, consistent PLS (PLSc, Dijkstra & Henseler, 2015a, 2015b) and integrated generalized structured component analysis (IGSCA, Hwang et al., 2021), which combines GSCA and GSCA with unique terms for accommodating measurement error (Hwang et al., 2017). Although these are great advancements, PLSc and IGSCA are not yet as mature as traditional SEM, which is highlighted by their limitations. Both PLSc and IGSCA lack a closed-form for the standard errors and must rely on bootstrap for statistical inference. Moreover, PLSc is rather limited when it comes to model specification. For instance, a researcher cannot impose constraints on the parameters (Tenenhaus, 2008) and is limited in specifying correlations among measurement errors (Rademaker et al., 2019). Similarly, IGSCA currently lacks a way to assess the overall model fit, which makes its application rather limited for explanatory modeling and theory testing (Hwang et al., 2021).

**Structural Equation Modeling With Latent, Emergent, and Excrescent Variables**

Only recently, Henseler (2021) sketched a specification that allows including composites in traditional SEM in a single step, while maintaining its flexibility, for example, specifying composites in an endogenous position. In this specification, the variables forming a composite are modeled by a set of synthetic variables, and the relationship between a composite and its observed variables is expressed by means of loadings instead of weights. The idea of specifying composites by means of loadings was previously proposed by Ogasawara (2007) in the context of canonical correlation analysis, where the relationships between canonical variates and the variables forming these variates are expressed by loadings (see also Gu et al., 2019 for an elaboration of this approach). Hence, in the following, I refer to this specification as the Henseler-Ogasawara (H-O) specification.

The H-O specification overcomes the shortcomings of the extant specifications known in SEM. Similar to the composite factor model specification, it circumvents the mismatch between the parameterization of composites and structural equation models by expressing composites in terms of loadings, not weights or structural coefficients. However, this specification of the composite model overcomes the identification issue of the composite factor model because it is more parsimonious (i.e., fewer parameters are required to describe the composite model, and accordingly, more degrees of freedom are available).

To present the H-O specification, I depart from the notion that the act of forming a composite represents a prescription for dimension reduction (Dijkstra & Henseler, 2011). However, in SEM, we are commonly less interested in a dimension reduction but rather interested in an understanding of how variables are related. Dimension reduction can be avoided if as many composites as exist in the dimensions of the vector space spanned by the set of observed variables are
built. This process is similar to a principal component analysis in which as many principal components as observed variables are extracted. As long as there is no perfect collinearity among the observed variables, that is, their variance-covariance matrix has full rank, the number of composites equals the number of observed variables.

Given that researchers are typically not interested in all possible composites but rather only one composite per set of observed variables, in the H-O specification, it is distinguished between two types of composites. Following the notion of Henseler (2021), I distinguish between emergent variables and excrescent variables. An emergent variable $\eta$ is a linear combination of variables that is related to other variables in the structural model; it represents a concept of scientific interest. Other names for emergent variables are “composite constructs” (Benitez et al., 2018) and “formative constructs”4 (Petter et al., 2007). In line with the composite model as studied in GSCA and PLS-PM (Cho & Choi, 2020; Dijkstra, 2017), the variables forming an emergent variable act along a single dimension—that is, they have conceptual unity, and all information between the variables forming an emergent variable and the other variables in the model is solely conveyed by the emergent variable. It is noted that this assumption may conflict with literature that does not ascribe conceptual unity to variables forming a composite (see e.g., Bollen & Bauldry, 2011; Bollen & Diamantopoulos, 2017) and composite specifications that allow for variables forming a composite to freely correlate with other variables not forming this composite (see e.g., Grace and Bollen, 2008). In contrast, this assumption aligns well with the literature that suggests to employ composites to model concepts that are formed and thus assumes that the construct emerge from its variables (e.g., Cole et al., 1993; Hancock, 1997; Henseler & Schuberth, 2021).

In contrast to an emergent variable, an excrescent variable $v$ is a linear combination of variables that is unrelated to all other variables in the structural model. Excrecent variables are required to capture the remaining variation in the set of observed variables and to ensure a correct parameterization but do not provide any other meaningful insight. Expressing one set of observed variables in terms of one emergent variable and several excrescent variables means that we replace Equation 1 by the following equation:

$$c = \begin{pmatrix} \eta \\ v \end{pmatrix} = W'x$$

Equation 3 essentially says that the vector of composites $c$, expressed as one emergent variable $\eta$ and one or more excrescent variables $v$, is a linear transformation of the observed variables $x$ using the square matrix $W$, which contains the weights.

As $W$ is a square matrix with a determinant different from zero, Equation 3 can be solved with regard to the observed variables, which leads to the following equation:

$$x = (W')^{-1} \begin{pmatrix} \eta \\ v \end{pmatrix} = \Lambda \begin{pmatrix} \eta \\ v \end{pmatrix}$$

As recognized by Ogasawara (2007) and shown in Equation 4, the inverse of the transposed weight matrix $W$ equals the matrix of the loadings $\Lambda$. Consequently, the relationships between the observed variables and the emergent and excrescent variables can be expressed by their covariances—that is, the composite loadings that are captured in the matrix $\Lambda$. Hence, the observed variables $x$ are a linear combination of the composites and composite loadings.

In contrast to the common factor model, composite models with emergent variables do not regard the observed variables as error-prone manifestations of a latent variable, but as dimensions of a vector space in which the emergent variable is located as one vector.

Figure 3 shows the specification for a simple model with one latent variable $\eta_1$ affecting the emergent variable $\eta_2$. While the latent variable is measured by the three observed variables $x_{11}$ to $x_{13}$, the emergent variable is formed by the three observed variables $x_{21}$ to $x_{23}$. To fully span the space of the observed variables forming the emergent variable, the emergent variable is accompanied by the two excrescent variables $v_{21}$ and $v_{22}$. Because both the emergent variable and the excrescent variables are composites, we symbolize them by means of hexagons, a shape introduced by Grace and Bollen (2008). In analogy to the distinction of shapes used for unobserved variables into larger circles (for latent variables) and smaller circles (for errors), I propose larger hexagons for emergent variables and smaller hexagons for excrescent variables.

### Identifying Structural Models Containing Composites

Analogous to any structural equation model, models including composites also need to be identified. In addition to the identification rules known from SEM and CFA, I focus on the identification rules that are specific to composites, that is, emergent and excrescent variables.

#### The Scaling Rule

Constructs have no natural scale; it is at the analyst’s discretion. The scaling rule thus entails that in some way, analysts need to provide a scaling for the constructs embedded in the model. This applies equally to common factors and composites. For composites, the scaling rule applies to both emergent and excrescent variables. There are two practices for accomplishing the scaling rule: Either a suitable composite loading is fixed to a predefined nonzero value, typically one, or the variance is fixed directly to a positive value, usually to one.

Fixing one observed variable’s loading is preferable over fixing the emergent variable’s variance for two reasons. First, this practice is also applicable to endogenous composites, whose variance cannot be constrained directly. Second, it avoids sign indeterminacy. Sign indeterminacy means that a set of parameters is identified with respect to their absolute values but not their sign. If the emergent variable’s variance is fixed, then $\lambda$ and $-\lambda$ are two vectors of loadings that yield the same model-implied variance-covariance matrix. In contrast, fixing a loading of an observed variable overcomes this problem.

#### Constraining the Loadings of Excrecent Variables

In addition to the emergent variables, composite models contain excrescent variables with their respective parameters. The task of the excrescent variables is to span the space of the observed variables forming the emergent variable jointly with the emergent variables

---

4 Note that we do not mean causal-formative measurement comprising latent variables, which does not refer to composites (Bollen & Bauldry, 2011). For a critical discussion, the interested reader is referred to Howell et al. (2007b); Bollen (2007); Bagozzi (2007); Howell et al. (2007a).
variable. In doing so, the space spanned only by the excrescent variables is orthogonal to the emergent variable. Because there is no particular interest in the excrescent variables, it does not matter how exactly they span the space as long as they are not perfectly collinear. Otherwise, the excrescent variables would not fully capture the observed variables' variance that could not be explained by the emergent variable. To ensure that the excrescent variables are not perfectly collinear, it is suggested to employ orthogonality constraints not only between the emergent variable and its excrescent variables but also among the excrescent variables themselves and between the excrescent variables and other variables in the model. Consequently, the excrescent variables are uncorrelated with all variables in the model except their associated observed variables.

To determine the parameters of the excrescent variables accompanying one emergent variable, the information contained in the variance-covariance matrix of the emergent variable's observed variables is exploited. However, not all moments are freely available because one moment is already exploited to determine a loading of the emergent variable. Consequently, if an emergent variable is formed by $J$ observed variables, there are $\frac{1}{2} J(J+1) - 1$ moments left to estimate the $J - 1$ variances and $J(J - 1)$ loadings of the excrescent variables, with $J - 1$ parameters overall.

Because there are more parameters than available moments, the balance must be settled by means of $\frac{1}{2} J(J - 1)$ constraints. To comply with the scaling rule, $J - 1$ parameters have to be fixed anyway. This leaves $\frac{1}{2} (J - 2)(J - 1)$ additional constraints to be set. Therefore, in the case of two observed variables (and thus one excrescent variable), no additional constraints have to be set; in the case of three observed variables (and thus two excrescent variables), one additional constraint is needed; in the case of four observed variables (and thus three excrescent variables), three additional constraints must be imposed; and so forth. A straightforward way of accomplishing this is to fix loadings of excrescent variables to zero in a cascading fashion: The second excrescent variable would have one loading fixed to zero, the third excrescent variable would have two loadings fixed to zero, and so forth.

**Estimating Structural Models Containing Composites**

In principle, all estimators known from SEM can be employed to estimate models containing composites. However, it cannot be ruled out that some of the current software implementations of SEM have been optimized toward common factor models and therefore (still) are suboptimal for estimation of models comprising composites. One indication pointing in that direction are the default starting values. Starting values for factor loadings appear suboptimal for composite models, particularly with regard to the excrescent variables for which loadings typically show alternating signs. It might be advisable to set the starting values of the free loadings to small values. Eventually, starting values closer to the final values would be helpful, for instance, generated by variance-based SEM (Bentler & Huang, 2014). Moreover, software implementations of SEM differ with regard to the optimizer employed and the maximum number of iterations, which might need to be adjusted.

As composites are specified via loadings instead of weights, current software implementations do not directly produce weight estimates. However, weight estimates can be obtained through inverting the matrix containing the respective loading estimates (see Equation 4). Consequently, current software implementations do not produce standard errors of the weight estimates. To address this issues, it was proposed to apply the delta method (Dorfman, 1938; Gu et al., 2019; Lu & Gu, 2018).
Applying the Henseler-Ogasawara Specification

In this section, I demonstrate how the H-O specification can be applied to specify composites in SEM. In the next subsection, SEM in combination with the H-O specification is applied to population variance-covariance matrices for different scenarios. This scenario analysis highlights the flexibility of the H-O specification and its advantage over the conventional model specification known from SEM. Concretely, it is shown that this specification is indeed capable of retrieving the parameters of different population models in the case of a correctly specified model. Furthermore, in the second subsection, I make use of a real dataset and illustrate that the H-O specification also functions under realistic conditions and not only in the ‘cozy’ environment of artificial data.

Artificial Data

To demonstrate the flexibility of the H-O specification, I make use of three scenarios. Concretely, I consider the three populations models as depicted in Table 1 to demonstrate that SEM applying the H-O specification is indeed able to retrieve the population parameters. All scenarios contain one composite and one common factor, but they are differently related to each other in the three scenarios. Scenario 1 shows a situation where the composite is in an exogenous position in the structural model. Although such a model can already be specified properly in SEM, it provides the opportunity to compare the results for the H-O specification to the conventional way of including composites in structural models, namely, the one-step approach. In Scenario 2, the composite and the common factor are not embedded into a structural model, that is, the two just correlate. In this case, the conventional SEM specification cannot be used because it does not allow specifying a correlation between a composite and a common factor since the composite is always treated as an endogenous variable. Consequently, one could only specify a correlation between the common factor and the disturbance term of the composite. Because the variance of the disturbance term is fixed to zero to ensure that the composite is fully composed by its observed variables, technically, the disturbance term cannot covary with other variables in the model. Moreover, the covariance between this disturbance term and the common factor is obviously not the same.

Table 1
Model Specification, Population Models, and the Variance-Covariance Matrix of the Three Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Model Specification</th>
<th>Population Model</th>
<th>Variance-covariance matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1a</td>
<td><img src="image" alt="Model Diagram" /></td>
<td><img src="image" alt="Population Model" /></td>
<td><img src="image" alt="Variance-Covariance Matrix" /></td>
</tr>
<tr>
<td>Scenario 1b</td>
<td><img src="image" alt="Model Diagram" /></td>
<td><img src="image" alt="Population Model" /></td>
<td><img src="image" alt="Variance-Covariance Matrix" /></td>
</tr>
<tr>
<td>Scenario 2</td>
<td><img src="image" alt="Model Diagram" /></td>
<td><img src="image" alt="Population Model" /></td>
<td><img src="image" alt="Variance-Covariance Matrix" /></td>
</tr>
</tbody>
</table>

Note. Random measurement errors are concealed in the model specifications.
as the covariance between the common factor and the composite. Finally, Scenario 3 is similar to Scenario 1, but the role of the composite and the common factor has changed, that is, the composite is now in an endogenous position in the structural model. Similar to Scenario 2, the conventional SEM specification to incorporate composites cannot be used. As the variance of the disturbance term of the composite is constrained to zero, such a model would imply that both the antecedents and the observed variables forming the composite together fully compose the composite. This is obviously not what is meant when placing a composite in an endogenous position.

Although the complexity of each model is kept simple, the three scenarios highlight the potential of the H-O specification because they cover all roles a composite can take in a classical linear structural model. In the following, I elaborate on each of the three scenarios.

**Scenario 1**

In Scenario 1, the population consists of the composite $\eta_1$ that is in an exogenous position in the structural model and predicts the common factor $\eta_2$. The composite is composed by three observed variables $x_1$ to $x_3$, while the common factor $\eta_2$ is associated with three indicators $y_1$ to $y_3$. The population model including the population parameter values and the corresponding variance-covariance matrix can be found in Table 1.

This scenario shows a situation for which both the H-O and the one-step specification can be applied in SEM to model the underlying population. Scenario 1a focuses on the H-O specification in which the block of the observed variables forming the composite is modeled by three variables: one emergent variable $\eta_1$ and two excretant variables $v_1$ and $v_2$. To identify the model, the third loading $\lambda_{13}$ of the emergent variable is fixed to one. The same is true for the second loading $\lambda_{22}$ of the first excretant variable $v_1$. Moreover, the first and the third loading of the second excretant variable $v_2$ are fixed to one and zero, respectively. Finally, the first factor loading $\lambda_{1}$ is fixed to one. The complete H-O specification is shown in Table 1 under “Model specification” for Scenario 1a.

Scenario 1b shows the conventional SEM specification, that is, the one-step approach, in which the observed variables composing the composite are elevated to the structural model, that is, the observed variables are antecedents of a latent variable for which disturbance term’s variance is fixed to zero. To identify the model, the path coefficient of the third composite’s observed variable is fixed to 1, that is, the third weight of the composite is fixed. The complete model specification is illustrated in Table 1 under “Model specification” for Scenario 1b.

**Scenario 2**

The population model of Scenario 2 is similar to the one of Scenario 1; however, the composite $\eta_1$ and the common factor $\eta_2$ are not embedded in a structural model in the population; the two just correlate. The values of the population parameters are the same as in Scenario 1, except that the path coefficient is replaced by a covariance. The population model including the values of the population parameters and the corresponding variance-covariance matrix are displayed in Table 1.

This scenario highlights the shortcomings of the one-step specification. Given that the observed variables forming the composite are elevated to the structural model, the composite appears as an endogenous variable. As a consequence, it is not possible to specify a covariance between the composite and other variables in the model, in this case, the common factor $\eta_2$. In contrast, the H-O specification allows specifying such a model. To identify the model, the same constraints as in Scenario 1 are imposed on the loadings of the observed variables associated with the common factor and the emergent and excretant variables. The H-O specification for Scenario 2 is illustrated in Table 1.

**Scenario 3**

Compared with Scenario 1, in the population of Scenario 3, the position of the composite and the common factor is reversed, that is, the composite appears as an endogenous variable in the structural model. It is emphasized that the composite in this population model should not be confused with a causal-formatively measured latent variable. Although the composite is an endogenous position in the structural model, it is still fully composed of its observed variables. For a discussion, the interested reader is referred to Aguirre-Urreta and Marakas (2014), Rigdon et al. (2014), Cadogan and Lee (2013), and Rigdon (2014). The population parameters are the same as in Scenario 1. The variance-covariance matrix of the observed variables and the population model including the values of the population parameters are displayed in Table 1.

This scenario further highlights the drawbacks of the conventional model specification to deal with composites in SEM. As the composite is in an endogenous position, it is not possible to fix the variance of the disturbance term to zero in order to ensure that the observed variables fully compose the composite; fixing the disturbance term’s variance to zero would imply that the observed variables forming the composite fully explain together with the common factor the variance of the composite. Obviously, this does not represent the population.

To retrieve the parameters of the specified models from the three scenarios, I use the maximum-likelihood estimator, as implemented in the open source R package lavaan (Rosseel, 2012; Version 0.6–8) and the commercial software Mplus (Muthén & Muthén, 1998–2017, Version 8). Both produce the same results. The syntax for the model specifications and the obtained results can be found in the online supplementary material. Table 2 shows the population parameters and the retrieved standardized parameters for the three scenarios.

As shown in Table 2, for Scenario 1, SEM applying the H-O specification (Scenario 1a) and the one-step approach (Scenario 1b) both retrieve the population parameters. Similarly, for Scenarios 2 and 3, the H-O specification can be used to obtain the population parameters when applied to the population variance-covariance matrix.

As discussed before, SEM based on the H-O specification does not directly produce the weights to form the composites—that is, the emergent and excretant variable—but the composite loadings—that is, the covariances between the observed variables and the composites. For the three considered scenarios, the produced standardized composite loadings, that is, the standardized parameters $\lambda_{11}$ to $\lambda_{13}$, are provided in Equation 5. It is noted that the (standardized) composite loadings are the same among the three considered scenarios because in the population model of each scenario the same weights and correlations among the observed variables forming the composite $\eta_1$ are assumed.
To ensure identifiability of the model, the weights of the single-indicator constructs are fixed to one. Moreover, the factor loading of $z_1$ is fixed to one. Considering the observed variables forming COMT ($z_{20}$ and $z_{21}$), the composite loading of $z_{20}$ on the exogenous variable $v_{12}$ and the composite loading of $z_{21}$ on the exogenous variable are fixed to one. Moreover, the composite loadings of the observed variables forming FKBp5 are constrained as follows:

$$
W = (\Lambda)^{-1} = \begin{pmatrix}
\eta_1 & v_1 & v_2 \\
0.400 & 0.533 & 0.745 \\
0.800 & 0.400 & -0.447 \\
0.800 & -0.600 & 0.000
\end{pmatrix}
$$

The first column of the matrix $W$ contains the standardized weights forming the emergent variable $\eta_1$ (highlighted in bold) as presented in Table 2 for the H-O specification. Additionally, the weights to form the exogenous variables $v_1$ and $v_2$ are obtained. However, they are usually of minor interest to a researcher.

**Empirical Example: The Gene and Depression Data of Hwang et al. (2021)**

To demonstrate how SEM based on the H-O specification can be applied to a real case, I consider an empirical example that was recently published in the study of Hwang et al. (2021) introducing IGSCA. In their study, the influence of genes on the severity of depression was examined based on a sample of 231 participants. As potential influences on the severity of depression, nine genes were considered. Moreover, it was controlled for the effects of gender ($z_{28b}$), age ($z_{25}$), and alcohol-related problems. To control for the latter, the Alcohol Use Disorders Identification test (AUDIT ($z_{28b}$), Saunders et al., 1993) was used. In contrast to the original study, I do not control for potential interaction effects. However, because the two interaction effects were not significant, I expect only minor differences with regard to the remaining parameter estimates. Figure 4 displays the structural model. Covariances among the exogenous constructs and the disturbance term are omitted to preserve readability.

Table 3 shows the constructs, their associated observed variables and how each construct was modeled, that is, as a common factor or composite. Severity of depression was measured by seven items of the HADS-D subscale (Oh et al., 1999) and modeled as a common factor. In contrast, the nine genes were modeled as composites where each gene is composed of a number of single nucleotide polymorphisms (SNPs). Specifically, the genes FKBp5 (Zobel et al., 2010) and COMT (Åberg et al., 2011) were composed of nine and two SNPs, respectively. The remaining seven genes, namely, SLC6A4 (Holmes et al., 2010), ADCYAP1R1 (Lowe et al., 2015), BDNF (Sen et al., 2003), HTR3A (Gatt et al., 2010), DRD2 (Vaske et al., 2009), NR3C1 (Galecka et al., 2013), and OXTR (McQuaid et al., 2013), were modeled as single-indicator composites. For an elaboration on the exact study design, the reader is referred to the original article of Hwang et al. (2021).

In the original study, IGSCA was chosen as an estimator because the proposed model contains both composites and a common factor. In the following, it is shown that this model can also be estimated by SEM applying the H-O specification. Figure 5 displays the H-O specification. To ensure readability, the covariances among the antecedents of depression are concealed. As can be seen in Figure 5, the observed variables forming FKBp5 were expressed by one emergent variable and eight excrescent variables ($v_{11}$ to $v_{18}$). Similarly, the two observed variables making up COMT were modeled by one emergent variable and one excrescent variable.

To ensure identification of the model, the weights of the single-indicator constructs are fixed to one. Moreover, the factor loading of $z_1$ is fixed to one. Considering the observed variables forming COMT ($z_{20}$ and $z_{21}$), the composite loading of $z_{20}$ on the exogenous variable $v_{12}$ and the composite loading of $z_{21}$ on the exogenous variable are fixed to one. Moreover, the composite loadings of the observed variables forming FKBp5 are constrained as follows:

$$
\begin{pmatrix}
\eta_1 & v_{11} & v_{12} & v_{13} & v_{14} & v_{15} & v_{16} & v_{17} & v_{18} \\
0.90 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.90 & \lambda_{10} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.90 & \lambda_{11} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.90 & \lambda_{12} & \lambda_{12,1} & 1 & 0 & 0 & 0 & 0 & 0 \\
0.90 & \lambda_{13} & \lambda_{13,1} & \lambda_{13,2} & 1 & 0 & 0 & 0 & 0 \\
0.90 & \lambda_{14} & \lambda_{14,1} & \lambda_{14,2} & \lambda_{14,3} & 1 & 0 & 0 & 0 \\
0.90 & \lambda_{15} & \lambda_{15,1} & \lambda_{15,2} & \lambda_{15,3} & \lambda_{15,4} & 1 & 0 & 0 \\
0.90 & \lambda_{16} & \lambda_{16,1} & \lambda_{16,2} & \lambda_{16,3} & \lambda_{16,4} & \lambda_{16,5} & 1 & 0 \\
0.90 & \lambda_{17} & \lambda_{17,1} & \lambda_{17,2} & \lambda_{17,3} & \lambda_{17,4} & \lambda_{17,5} & \lambda_{17,6} & 1 \\
\end{pmatrix}
$$

To ensure identification, for each excrescent one loading is constrained to one to fix the excrescent variables’ scale. Moreover, the loadings of the excrescent variables are fixed to zero in a cascading fashion; that is, for the second excrescent variable $v_{12}$ one loading is fixed to zero, for the third excrescent variable $v_{13}$ two loadings are fixed to zero and so forth.

To obtain the results for SEM applying the H-O specification, the maximum-likelihood estimator as implemented in Mplus (Muthén & Muthén, 1998–2017; Version 8) and lavaan (Rosseel, 2012) was used. In contrast to the original study, I do not control for potential interaction effects. However, because the two interaction effects were not significant, I expect only minor differences with regard to the remaining parameter estimates. Figure 4 displays the structural model. Covariances among the exogenous constructs and the disturbance term are omitted to preserve readability.
2012; Version 0.6–8) was used. The variance-covariance of the observed variables reported in the original study (see Table 7 in Hwang et al., 2021), served as input for the data analysis (see Table 5 in the online supplementary material). The model specification for lavaan and Mplus can be found in the online supplementary material. Because the results between Mplus and lavaan do not differ, in the following only the Mplus results are reported. The complete results can be found in the online supplementary material.

Table 4 contrasts the standardized parameter estimates for SEM based on the H-O specification with those from IGSCA and PLSc. Because the loadings and weights for single-indicator constructs were fixed to one, they are omitted from the table. Considering the H-O specification, the weight estimates were calculated based on Equation 4; that is, they were obtained from the estimated loadings. Although, the standard errors for the weight estimates can be generally obtained via the delta method (Gu et al., 2019; Lu & Gu, 2018), they are not reported in the following. The IGSCA results were copied from the original study of Hwang et al. (2021; see Tables 6 and 7 of their study). It is noted that in their model, although not significant, two interaction terms were included. Moreover, weight estimates of IGSCA are not included as they were not reported in the original study. To obtain the PLSc estimates, the R package cSEM (Rademaker & Schuberth, 2020) was used. For composites, Mode B was used to obtain the weight estimates, while for parameters related to the common factor, Mode A in combination with a correction for attenuation was applied. For the inner weighting, the path weighting scheme was employed. Given that the original dataset was not available, no standard errors could be computed for the PLSc estimate because for PLSc, no closed-form standard errors exist and bootstrap is used for their estimation. Consequently, no inference is made for PLSc estimates. The model specification and the complete results can be found in the online supplementary material.

The results for the three approaches are quite similar for most of the parameters. The estimated factor loadings and path coefficients show similar sizes, signs, and significances. It is noted that the two moderation effects are only included in IGSCA, and thus, their results are not available for SEM and PLSc. Considering the two composites FKBP5 and COMT, the composite loading estimates show substantial differences among SEM, IGSCA, and PLSc. While IGSCA produced loadings that were all above .8 and significant, in SEM, most of the loading estimates are rather small and insignificant. Similarly, the estimated composite loadings of PLSc

5 The results were additionally verified by ADANCO (Henseler & Dijkstra, 2015).
are much smaller than the corresponding estimates of IGSCA. In general, the loadings of PLSc are more similar to those produced by SEM than those produced by IGSCA. In addition to the composite loadings, the weights are displayed for SEM and PLSc. While the weights are automatically estimated by the iterative PLS algorithm for PLSc, the standardized composite loading estimates are transformed to obtain the standardized weight estimates for SEM. For IGSCA, the weights are not reported as they were not presented in the original study. The weight estimates of SEM and IGSCA are compared to determine which method provides more accurate estimates. In addition to the composites in the structural model, the exogenous variables in the model are also correlated with other variables. Consequently, the presented H-O specification overcomes the drawbacks of the existing model specification in addressing composites.

Researchers often distinguish between factor-based and composite-based SEM to highlight whether traditional SEM or approaches such as PLS or GSCA are used (e.g., Hair et al., 2017; Hwang et al., 2021; Rigdon, 2012). The presented H-O specification contributes to overcoming this dichotomy and further contributes to unifying the two worlds. While recent developments such as IGSCA and PLSc have opened composite-based SEM to estimating models containing common factors, the study at hand enables factor-based SEM to cope with composites. Because the line between composite-based and factor-based SEM has become increasingly blurred, I recommend distinguishing between the model and the estimator employed instead of differentiating types of SEM.

The H-O specification not only enhances SEM by allowing modeling with composites but also renders all the developments of SEM accessible to composite modeling that are currently not available if approaches such as PLS(c) or (I)GSCA are employed. For instance, the $\chi^2$ test and various fit indices can be used to judge the overall model fit. In analogy to models containing common factors, it is likely that the model fit reduces the larger the number of composites comprised in a model because more constraints are imposed on the variance-covariance matrix of the observed variables that likely do not fully hold in empirical settings. In addition to the overall model fit assessment, SEM provides more sophisticated ways of treating missing values compared to PLS(c) and (I)GSCA (Allison, 2003). Moreover, the H-O specification allows that recent developments in the context of composite modeling such as confirmatory composite analysis (CCA, Schuberth et al., 2018; Henseler & Schuberth, 2020) can be executed using common SEM software such as Mplus or the R package lavaan. Finally, researchers can choose from various estimators available for SEM to estimate their models containing composites. However, during the estimation of the models studied in this article, it turned out

6In their original article on PLSc, Dijkstra and Henseler (2015a) mention the possibility of deriving such a test but propose a bootstrap-based test to assess the overall model fit. Since the original sample is not available, the bootstrap-based test for overall model fit could not be conducted.

7Similarly, the literature distinguishes between variance-based and covariance-based SEM (e.g., Henseler, 2017).

<table>
<thead>
<tr>
<th>Observed variable</th>
<th>Construct</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>z_1: HADS2</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_2: HADS5</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_3: HADS6</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_4: HADS8</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_5: HADS10</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_6: HADS12</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_7: HADS14</td>
<td>Depression</td>
<td>Common factor</td>
</tr>
<tr>
<td>z_8: rs25531</td>
<td>SLC6A4</td>
<td>Composite</td>
</tr>
<tr>
<td>z_9: rs9296158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{10}: rs3800373</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{11}: rs1360780</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{12}: rs9470080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{13}: rs4713916</td>
<td>FKB5</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{14}: rs4713919</td>
<td>FKB5</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{15}: rs6902321</td>
<td>COMT</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{16}: rs56311918</td>
<td>ADCYAP1R1</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{17}: rs3798345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{18}: rs2267735</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{19}: rs6265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{20}: rs4680</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{21}: rs4633</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{22}: rs1062613</td>
<td>HTR3A</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{23}: rs2075652</td>
<td>DRD2</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{24}: rs258747</td>
<td>NR3C1</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{25}: rs53576</td>
<td>OXTR</td>
<td>Composite</td>
</tr>
<tr>
<td>z_{26}: Gender</td>
<td></td>
<td>Covariate</td>
</tr>
<tr>
<td>z_{27}: Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z_{28}: AUDIT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5
Henseler-Ogasawara Specification for the Empirical Example
that the default starting values for the loadings of the emergent and excrescent variables are suboptimal—that is, they often lead to convergence problems. These findings may indicate that algorithms have been optimized for common factor models rather than composite models expressed by emergent and excrescent variables. Using variance-based SEM to generate starting values might be a viable option (Bentler & Huang, 2014).

The empirical example showed that most of the parameter estimates are in line among SEM, IGSCA, and PLSc. However, with regard to the composite loadings and the weights forming the composites, substantial differences could be observed. This is surprising because although the maximum-likelihood estimator used in SEM and IGSCA are full-information estimators and PLSc is a limited information estimator, all three approaches produce consistent estimates when the model is correctly specified. Consequently, all three estimators are expected to produce similar results. Future research is therefore tasked with shedding light on the differences of the various approaches and highlighting situations in which one approach is

Table 4
Comparison of the Standardized Parameter Estimates Across SEM, IGSCA, and PLSc

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SEM (H-O)</th>
<th>IGSCA</th>
<th>PLSc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
</tr>
<tr>
<td>Loadings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.67*</td>
<td>0.04</td>
<td>0.68*</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.75*</td>
<td>0.04</td>
<td>0.74*</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>0.82*</td>
<td>0.03</td>
<td>0.81*</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.49*</td>
<td>0.06</td>
<td>0.51*</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>0.65*</td>
<td>0.04</td>
<td>0.66*</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>0.76*</td>
<td>0.03</td>
<td>0.76*</td>
</tr>
<tr>
<td>$\lambda_7$</td>
<td>0.55*</td>
<td>0.05</td>
<td>0.56*</td>
</tr>
<tr>
<td>Loadings (weights)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBP5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{10}$ ($w_{10}$)</td>
<td>0.01 (0.31)</td>
<td>0.18</td>
<td>0.89*</td>
</tr>
<tr>
<td>$\lambda_{11}$ ($w_{11}$)</td>
<td>0.02 (1.36)</td>
<td>0.18</td>
<td>0.89*</td>
</tr>
<tr>
<td>$\lambda_{12}$ ($w_{12}$)</td>
<td>0.24 (0.33)</td>
<td>0.18</td>
<td>0.98 (1.92)</td>
</tr>
<tr>
<td>$\lambda_{13}$ ($w_{13}$)</td>
<td>0.11 (0.37)</td>
<td>0.18</td>
<td>0.91*</td>
</tr>
<tr>
<td>$\lambda_{14}$ ($w_{14}$)</td>
<td>0.56* (1.19)</td>
<td>0.15</td>
<td>0.86*</td>
</tr>
<tr>
<td>$\lambda_{15}$ ($w_{15}$)</td>
<td>0.09 (0.65)</td>
<td>0.18</td>
<td>0.89*</td>
</tr>
<tr>
<td>$\lambda_{16}$ ($w_{16}$)</td>
<td>0.32 (0.14)</td>
<td>0.17</td>
<td>0.83*</td>
</tr>
<tr>
<td>$\lambda_{17}$ ($w_{17}$)</td>
<td>0.17 (0.36)</td>
<td>0.18</td>
<td>0.90*</td>
</tr>
<tr>
<td>COMT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{20}$ ($w_{20}$)</td>
<td>0.03 (2.51)</td>
<td>0.19</td>
<td>0.98*</td>
</tr>
<tr>
<td>$\lambda_{21}$ ($w_{21}$)</td>
<td>0.39* (2.72)</td>
<td>0.17</td>
<td>0.98*</td>
</tr>
<tr>
<td>Path coefficients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC6A4 → Dep.</td>
<td>–0.01</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>FKBP5 → Dep.</td>
<td>–0.05</td>
<td>0.07</td>
<td>–0.06</td>
</tr>
<tr>
<td>ADCYAP1R1 → Dep.</td>
<td>0.18*</td>
<td>0.07</td>
<td>0.18*</td>
</tr>
<tr>
<td>BDNF → Dep.</td>
<td>–0.08</td>
<td>0.07</td>
<td>–0.05</td>
</tr>
<tr>
<td>COMT → Dep.</td>
<td>–0.13</td>
<td>0.07</td>
<td>–0.03</td>
</tr>
<tr>
<td>HTR3A → Dep.</td>
<td>0.11</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>DRD2 → Dep.</td>
<td>–0.08</td>
<td>0.07</td>
<td>–0.10</td>
</tr>
<tr>
<td>NR3C1 → Dep.</td>
<td>0.03</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>OXTR → Dep.</td>
<td>–0.06</td>
<td>0.07</td>
<td>–0.07</td>
</tr>
<tr>
<td>Gender → Dep.</td>
<td>–0.23*</td>
<td>0.08</td>
<td>–0.24*</td>
</tr>
<tr>
<td>Age → Dep.</td>
<td>–0.04</td>
<td>0.07</td>
<td>–0.06</td>
</tr>
<tr>
<td>AUDIT → Dep.</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Gender × Age → Dep.</td>
<td>NA</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>Gender × AUDIT → Dep.</td>
<td>NA</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>Fit statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$ statistic</td>
<td>277.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p$-value (df)</td>
<td>0.09 (247)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLI</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRMR</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. SEM = structural equation modeling; IGSCA = integrated generalized structured component analysis; PLSc = consistent partial least squares; CFI = comparative fit index; TLI = Tucker–Lewis index; SRMR = standardized root mean square residual; AUDIT = Alcohol Use Disorders Identification test; NA = not available. * = estimate is significant on a 5% level.
preferred over the others. Moreover, in the study at hand, composites were composed of observed variables that are assumed to be free from measurement error, which is likely not given in empirical settings. Hence, future research might investigate specifications that relax this assumption, for example, specifying composite made up of common factors. Such a specification would be particularly valuable for modeling aggregate constructs (e.g., Edwards, 2001). Moreover, the standard error estimates for the weights needed to be derived. A viable solution might be the delta method.

Finally, composites are often used to approximate common factors. It is well known in the literature that this approach leads to inconsistent estimated parameters if the common factor model is true because of the attenuation bias (Cohen et al., 1990). However, recent literature has shown that under alternative measurement structure—that is, in situations where the common factor is not true—the bias can be much more variable and often more severe if a common factor model is mistakenly fitted compared to first creating composite scores and subsequently estimating the parameters of the structural model (Rhemtulla et al., 2020). Although the H-O specification might provide a middle way between using sum scores and estimating a common factor model in terms of freely estimated parameters, future research is asked to investigate whether the H-O specification shows benefits over the use of “simple” sum scores to approximate common factors or alternative measurement structures.

References


