Meta-analytic structural equation modeling with moderating effects on SEM parameters

Jak, S.; Cheung, M.W.-L.

Meta-Analytic Structural Equation Modeling With Moderating Effects on SEM Parameters

Suzanne Jak
University of Amsterdam

Mike W.-L. Cheung
National University of Singapore

Abstract

Meta-analytic structural equation modeling (MASEM) is an increasingly popular meta-analytic technique that combines the strengths of meta-analysis and structural equation modeling. MASEM facilitates the evaluation of complete theoretical models (e.g., path models or factor analytic models), accounts for sampling covariance between effect sizes, and provides measures of overall fit of the hypothesized model on meta-analytic data. We propose a novel MASEM method, one-stage MASEM, which is better suitable to explain study-level heterogeneity than existing methods. One-stage MASEM allows researchers to incorporate continuous or categorical moderators into the MASEM, in which any parameter in the structural equation model (e.g., path coefficients and factor loadings) can be modeled by the moderator variable, while the method does not require complete data for the primary studies included in the meta-analysis. We illustrate the new method on two real data sets, evaluate its empirical performance via a computer simulation study, and provide user-friendly R-functions and annotated syntax to assist researchers in applying one-stage MASEM. We close the article by presenting several future research directions.

Translational Abstract

Meta-analytic structural equation modeling (MASEM) is an increasingly popular statistical technique that combines the strengths of meta-analysis and structural equation modeling. Meta-analysis is useful in combining results from different studies, whereas structural equation modeling allows researchers to test different theoretical models. MASEM facilitates the evaluation of complete theoretical models and tests how good the proposed models fit the published data. Because published studies may be different in terms of samples and measurements, the findings are likely heterogeneous, that is, nonidentical. We propose a novel MASEM method, one-stage MASEM, which is better suitable to explain study-level heterogeneity than existing methods. One-stage MASEM allows researchers to use continuous or categorical moderators, for example, the mean age of the participants and gender ratio, to explain the differences across studies with potential missing data. We illustrate the new method on two real data sets, and provide user-friendly R-functions and annotated syntax to assist researchers in applying one-stage MASEM. We close the paper by presenting several future research directions.

Keywords: meta-analytic structural equation modeling, meta-analysis, structural equation modeling, moderation analysis, heterogeneity

Supplemental materials: http://dx.doi.org/10.1037/met0000245.supp

With a growing number of empirical studies being published every year, there is an increasing need to systematically review the available empirical evidence. Consequently, meta-analysis is rapidly gaining popularity in various scientific fields, including psychology (Schulze, 2007). Combining data from separate independent studies, meta-analysis allows researchers to estimate an average effect size, estimate the degree of heterogeneity across studies, and to explain the heterogeneity across studies using moderator variables.

Originally, meta-analysis was developed to analyze bivariate effect sizes, representing the relationship between only two vari-

ables (Glass, 1976). There are several advances extending meta-analysis to address more complicated research questions. These include, to name a few, analyzing more than one effect size with a multivariate meta-analysis (e.g., Becker, 2007; Cheung, 2013; Gleser & Olkin, 2009; Jackson, Riley, & White, 2011), handling nonindependent effect sizes with either a three-level meta-analysis (e.g., Cheung, 2014b; Konstantopoulos, 2011; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013) or robust standard errors (Hedges, Tipton, & Johnson, 2010; Tipton, 2015; Tipton & Pustejovsky, 2015), meta-analyzing partial effect sizes (Aloe, 2014; Aloe & Becker, 2012; Becker & Wu, 2007), and conducting meta-analysis within the Bayesian framework (e.g., Schmid, 2001; Sutton & Abrams, 2001). These advances make meta-analysis a powerful tool to synthesize existing research findings and to help resolve the replicability crisis in psychology (Anderson & Maxwell, 2016; Hedges & Schauer, 2018).

Another line of extension of standard meta-analysis is meta-analytic structural equation modeling (MASEM). MASEM is a statistical technique to fit hypothesized models on sets of variables from multiple independent studies (Becker, 1992, 1995; Cheung, 2015a, 2019; Jak, 2015; Viswesvaran & Ones, 1995). MASEM combines the strengths of meta-analysis (systematic synthesis of study-results) and structural equation modeling (SEM; fitting models with intricate relations between observed and latent variables). A MASEM analysis typically consists of two stages (Viswesvaran & Ones, 1995). In Stage 1, correlation matrices from different studies are combined to form a pooled correlation matrix. In Stage 2, a structural equation model, such as a path model or factor analytic model, is fitted to the pooled correlation matrix.

The study of Topa and Moriano (2010) illustrates how MASEM is used to answer research questions. Topa and Moriano (2010) gathered correlation coefficients between social norms, attitudes, perceived control, smoking intention, and smoking from 35 published studies, obtaining a total of 217 correlations. A synthesis of the observed correlations across studies provided a pooled correlation matrix of the five variables. By fitting the path model from Figure 1a to the pooled correlation matrix, they tested how well the theory of planned behavior could predict smoking behavior. Among other results, they found that social norms better predicted smoking intentions than attitudes and perceived behavioral control. If these authors had not applied MASEM, they might have erroneously concluded that behavioral intention is most predictive of the intention to smoke because this variable showed the largest pooled bivariate correlation with smoking intention.

MASEM thus allows researchers to evaluate the unique effects of multiple predictors simultaneously. To give another example of MASEM, Zeegers, Colonnesi, Stams, and Meins (2017) used MASEM on the data from 63 independent studies to test the model in Figure 1b. These researchers were interested in the indirect effect of parent's mentalization (defined as the degree to which parents show frequent, coherent, or appropriate appreciation of their infants' internal states) on parent–child attachment via parent's sensitivity. They found that there was a significant small positive indirect effect, as well as a significant positive medium sized direct effect, representing partial mediation of the effect of mentalization on attachment via sensitivity.

MASEM is most commonly applied to evaluate path models, such as in the two examples provided above, but the technique can also be used for models with latent variables. Hong and Cheung (2015), for example, obtained 159 correlation coefficients between six cognitive vulnerabilities associated with depression from 73 studies. They applied meta-analytic confirmatory factor analysis to evaluate two alternative factor analytic models underlying the associations between the vulnerability measures. A one-factor model provided the best fit to the meta-analytic data (see Figure 1c).

In the example of Topa and Moriano (2010), the authors observed 217 correlations, while there would be 35 (studies) × (5 × 4)/2 (correlations) = 350 correlations observed if all studies provided complete data. Zeegers et al. (2017) observed 88 correlations, while there would be 63 × (3 × 2)/2 = 189 correlations observed if all studies provided complete data. Hong and Cheung (2015) gathered 159 coefficients, while complete data would yield 73 × (6 × 5)/2 = 1,095 observed correlation coefficients. These examples illustrate one of the strengths of MASEM; it can test new theories that possibly have not been tested in any of the primary studies (Bergh et al., 2016; Viswesvaran & Ones, 1995). In other words, the primary studies may have only tested some of the variables included in the MASEM models. For a MASEM, the requirement is that all the bivariate correlations among these five variables are present in the average correlation matrix, but they may be missing in part of the primary studies (Jak & Cheung, 2018a).

However, there is an important caveat in fitting MASEM on the bivariate correlations based on different studies. Let us use the theory of planned behavior in Figure 1a as an example. Suppose that the average correlation coefficients between social norms, attitudes, and perceived control are based on the children population, whereas the average correlation coefficient between smoking intention and smoking behavior is from the adult population. It does not make any theoretical sense to combine these correlation coefficients from different populations into an average correlation matrix for MASEM. Therefore, researchers should set clear inclusion criteria based on what studies can be meaningfully combined before conducting a MASEM (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011).

While standard meta-analysis evaluates each relationship between the variables of interest in a univariate way, MASEM is a multivariate technique that evaluates complete theoretical models, accounts for sampling covariance between effect sizes, provides the researcher measures of overall fit of a hypothesized model, and provides parameter estimates from SEMs with confidence intervals and standard errors. It is for these reasons that MASEM is increasingly applied in the social sciences (Sheng, Kong, Cortina, & Hou, 2016).

We summarize some key procedures and decisions that researchers have to make when conducting a MASEM (see Cheung, 2015a; Viswesvaran & Ones, 1995; also see Cooper, 2010 for some general issues in meta-analysis):

1. Identify key constructs, measurements, and structural equation models. Before conducting the meta-analysis, researchers have to identify all the relevant key constructs, measurement models, and structural equation models. Relatively smaller models, with not too many observed variables, are generally better analyzable than large models. Therefore, it is advisable to be as strict as the theory permits with the identification of relevant

a

b

c
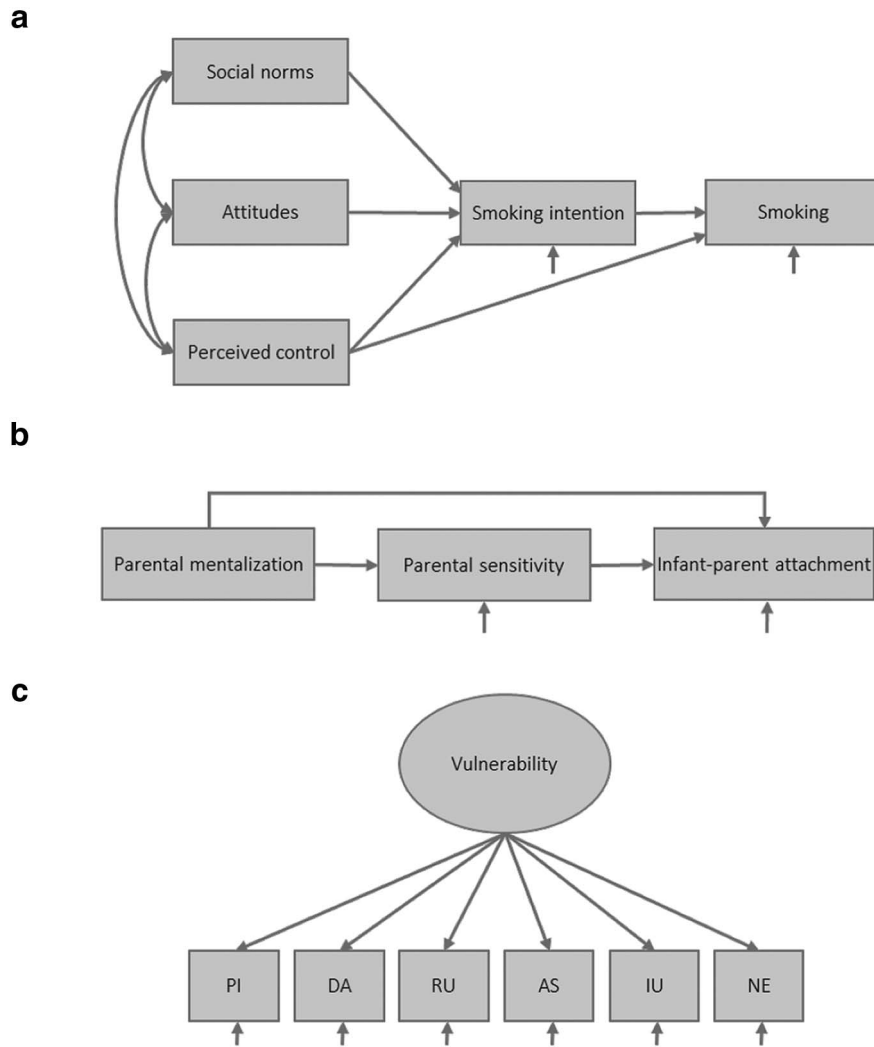
Figure 1. Three models from applications of MASEM. In panel c: PI = pessimistic inferential style; DA = dysfunctional attitudes; RU = ruminative style; AS = anxiety sensitivity; IU = intolerance of uncertainty; NE = fear of negative evaluation.

constructs. Researchers may formulate several theoretically meaningful models for future testing.

2. Formulate inclusion/exclusion criteria. This step is essential as it provides theoretical justifications on whether the selected studies can be meaningfully combined. For example, researchers may need to ask whether it makes sense to combine studies from different populations such as children versus adults and studies that use different measurement instruments.

3. Identify and extract all the relevant correlation matrices, sample sizes, and study characteristics (moderators). The basic inputs in a MASEM are the correlation matrices and their sample sizes. It is quite likely that some of the correlations are incomplete (please refer to the later sections on how missing data are handled in MASEM). Researchers should attempt to retrieve additional unpub-

lished data, such as dissertations and conference presentations, in order to minimize the influence of publication bias. Also, it is important that the correlation matrices are assumed independent. That is, each study only contributes one correlation matrix. When the correlation matrices are not independent, either a multilevel model (Wilson, Polanin, & Lipsey, 2016) or a robust standard error approach (Hedges et al., 2010; Tipton, 2015; Tipton & Pustejovsky, 2015) may be used to handle the dependency in the data.

4. Choose an appropriate approach to combine the correlation matrices. Several models, for example, the univariate-r (Viswesvaran & Ones, 1995), generalized least squares (GLS; Becker, 1992, 1995), or two-stage SEM (TSSEM; Cheung, 2014a; Cheung & Chan, 2005), have been proposed in the literature. The univariate-r

approach independently meta-analyzes the cells in the correlation matrix, whereas the GLS and TSSEM approaches take the dependence of the correlations in the correlation matrix into account with a multivariate approach.

5. Fit the structural equation models with the associated approach. If the univariate-r approach is used in the above Step 4, the average correlation matrix is treated as if it was an observed covariance matrix in fitting the structural equation models. A harmonic mean of the sample sizes is usually used as the sample size in SEM. If the GLS approach is used, the parameters and their standard errors in the path models are estimated using matrix operations and the multivariate delta method (see Becker, 1992 for the technical details). The TSSEM approach uses weighted least squares (WLS) estimation, also known as the asymptotically distribution-free (ADF) in SEM, to weigh the precision of the average correlation matrix in fitting the structural equation models. Readers may refer to Cheung (2015a) for the details on the differences between these approaches.

Notwithstanding the benefits of MASEM, the current MASEM-methods do not take full advantage of a major strength of standard meta-analysis, which is the ability to explain heterogeneity in effect sizes across studies by including study-level moderators. Typical moderators in meta-analyses are, for example, the mean age of participants and the percentage of males in the sample. Some existing MASEM methods are able to account for heterogeneity, thereby providing unbiased parameter estimates and significance tests (Cheung & Jak, 2017), but are not well suited to test the effect of continuous study-level variables on the parameters in the structural models. Other MASEM methods allow for the inclusion of continuous moderators but do not account for missing correlations in the primary studies (Cheung & Cheung, 2016). Therefore, all three applications discussed before used standard univariate metaregression in addition to MASEM to evaluate whether variability in the study's individual correlation coefficients could be explained by continuous study-level moderators.

For example, Zeegers et al. (2017) fitted the path model from Figure 1b to evaluate the hypothesized mediational effect of parent's mentalization on parent–child attachment through parental sensitivity. They were also interested in whether the effects of mentalization and sensitivity were moderated by the children's age at the time of assessment. However, because it was not possible to include age as a moderator in the path model, they tested the moderation effects on the bivariate correlations between the variables in three separate analyses. That is, they performed a univariate random-effects meta-analysis on the correlation coefficients between sensitivity and attachment and then included the average age of children in the sample as a moderator variable in a metaregression to explain study-level heterogeneity on the bivariate correlations. Next, they used this same procedure to test moderator effects on the correlation between sensitivity and mentalization, and the correlation between mentalization and attachment.

Although there are well-established techniques to test moderation in multiple regression (Aiken, West, & Reno, 1991) and SEM

(e.g., Klein & Muthén, 2007; Lee & Zhu, 2002; Little, Bovaird, & Widaman, 2006; Marsh, Wen, & Hau, 2004), none of these techniques can directly be applied to MASEM. In most SEM applications, except for multilevel SEM, researchers model the *within*-study variations. However, when only the summary statistics (correlation coefficients) are available for a research synthesis, researchers can only model the *between*-study variations in the meta-analysis or MASEM (Shadish, 1996).

Aim of This Study

It would be highly desirable if fitting the SEM and evaluating heterogeneity could be performed within the same framework. Moreover, researchers applying MASEM are presumably more interested in testing whether the effects in the hypothesized SEM-model (such as regression coefficients in a path model and factor loadings in a factor model) are moderated, than in the evaluation of moderating effects on the bivariate correlations. The purpose of this article is, therefore, to present a novel approach that allows researchers to incorporate moderator variables into their MASEM, in which all the parameter in the SEM can be modeled by the moderator variable. The new random-effects MASEM approach does not involve two strictly separated stages. Therefore, we call it "one-stage MASEM."

Before we present one-stage MASEM in detail, we will review the currently available MASEM methods. First, we illustrate the fixed- and random-effects models in MASEM. Then we discuss correlation- and parameter-based MASEM, including full information MASEM and Bayesian MASEM. Next, we will introduce one-stage MASEM, illustrate its use on two real data sets, and present a simulation study to evaluate the empirical performance.

Fixed- and Random-Effects Models

Similar to standard (univariate) meta-analysis, MASEM consists of fixed- and random effects methods. Fixed-effects methods (or common effects models) assume that all studies share the same population effect sizes, so that observed differences between effect sizes are the result of sampling error (Hedges & Vevea, 1998). The primary goal of the fixed-effects model is to summarize the studies in a meta-analysis. Random effects models, on the other hand, assume a "super distribution" of population effect sizes. That is, each study is assumed to have its own population effect size. The mean (often denoted with the symbol ρ) and variance (often denoted with the symbol τ^2) of the "super distribution" of population effect sizes is estimated in the random-effects analysis, while also accounting for sampling error in the individual studies (sampling variance is often represented by v). The primary goal of a random-effects model is to generalize findings of a meta-analysis beyond the studies included in the meta-analysis.

Mathematically, the fixed-effects model is a special case of the random-effects model by setting $\tau^2 = 0$. If the fixed-effects model is incorrectly applied to data from a random-effects model, the estimated confidence intervals tend to be too short (Schmidt, Oh, & Hayes, 2009). Because it is very likely that the population effect sizes differ across studies, for example, because different studies focus on slightly different populations, or use different measurement instruments, the random-effects model is generally preferred (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hedges, 2016).

In the remainder of the article, we focus on random-effects models unless it is explicitly stated that we are discussing a fixed-effects model.

When the variance component gets larger, the usefulness of the average effect is questionable (Kisamore & Brannick, 2008), because the population effects may vary a lot. Let us consider a simple example with $\rho = .1$ and $\tau = .2$. Assuming that the population effects sizes are normally distributed, about 95% of the population effect sizes lay between $.1 \pm 1.96 \times .2$ or $(-.29, .49)$. This indicates that the population effect size can be either very positively large (.49) or negative ($-.29$). Similar concerns have been raised in MASEM. That is, even if the proposed structural equation model fits well on the average correlation matrix, it does not necessarily mean that the proposed model fits equally well in all population correlation matrices (see Cheung & Cheung, 2016 for the discussion). The proposed one-stage MASEM in this study resolves this issue by allowing studies with different parameter estimates depending on the values of the moderator.

Correlation- and Parameter-Based MASEM

Cheung and Cheung (2016) differentiated two types of random-effects MASEM: correlation-based and parameter-based MASEM.[1] In correlation-based MASEM, the correlation matrices are the effect sizes and source of heterogeneity in the meta-analysis. On the other hand, parameter-based MASEM considers the estimated parameters in SEM as the effect sizes and source of heterogeneity in the meta-analysis. We will discuss the current methods using this classification, and also we discuss one other method that is not specifically correlation or parameter-based MASEM. For each method, we indicated in Table 1 whether the method provides estimates of the heterogeneity, if the method allows the inclusion of continuous moderator variables to explain heterogeneity and if the method is applicable with missing data.

Correlation-Based MASEM

Two-stage methods for correlation-based MASEM have in common that there is a first stage, in which a pooled correlation matrix is estimated using fixed- or random effects meta-analysis, and a second stage in which the parameters of the structural equation model are estimated. Becker (1992, 1995) proposed using GLS to estimate a pooled correlation matrix and its asymptotic sampling covariance matrix. Next, the pooled correlation matrix and its asymptotic sampling covariance matrix are used to fit path models. Viswesvaran and Ones (1995) proposed applying univariate meta-analyses on the individual correlations to construct a pooled correlation matrix, and using the pooled matrix as if it was an observed covariance matrix to fit a structural equation model. For the ease of reference, we call this approach the univariate-r approach here. Based on the work of Becker (1992), Cheung and Chan (2005; Cheung, 2014a) proposed the TSSEM approach to conduct MASEM for path models, confirmatory factor analytic models, and SEM. Figure 2 provides a conceptual overview of the two stages of TSSEM.

Because in two-stage correlation-based MASEM the structural equation model is fitted on the pooled correlation matrix, and not directly on the observed correlations, it is not possible to evaluate the heterogeneity of parameters from the structural equation model from Stage 2 (see rows 1–3 of Table 1). One solution is creating

subgroups of studies based on the moderator variable of interest and comparing the estimates in the subgroups. Categorizing continuous moderator variables, however, leads to a loss of information on the moderating variable, and always results in smaller numbers of studies to be analyzed within each group, with associated estimation problems (Jak & Cheung, 2018b; MacCallum, Zhang, Preacher, & Rucker, 2002). Note that at Stage 2 of correlation-based MASEM, it *is* possible to evaluate heterogeneity of the correlation coefficients, using multivariate metaregression (Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, & Colditz, 1998; Jackson, White, & Thompson, 2010).

Yu, Downes, Carter, and O'Boyle (2016) proposed a way of quantifying the heterogeneity at Stage 2 in terms of credible intervals of SEM-parameters by fitting structural equation models from correlation matrices generated from a parametric bootstrap. They called this method Full Information MASEM (FIMASEM). FIMASEM involves three steps: (a) estimating the pooled correlations and their variances, (b) performing a parametric bootstrap to generate a large number of heterogeneous correlation matrices, and (c) fitting the SEM to all generated matrices. Although this method provides correct credible intervals, the obtained test statistics and goodness-of-fit indices are questionable, making it impossible to evaluate the fit of the hypothesized models with this approach (Cheung, 2018; cf. Yu, Downes, Carter, & O'Boyle, 2018). Moreover, because Yu et al. (2016) did not discuss how moderators could be handled in their approach, it is still unclear how their approach can be extended to models with moderators (see row 4 of Table 1). Because the correlation matrices that are generated with the parametric bootstrap are not tied to specific studies and their associated values on the moderator variables, it seems impossible to explain between-studies variability in SEM parameters with continuous moderator variables (Ke, Zhang, & Tong, 2018).

Parameter-Based MASEM

An alternative way of evaluating SEM on meta-analytic data is to fit the SEM in each of the primary studies, and then combine the parameters of interest (e.g., factor loadings or path coefficients) as effect sizes in a meta-analysis. This method (called "parameter-based MASEM" by Cheung & Cheung, 2016) has indeed been proposed to combine regression coefficients (Becker & Wu, 2007) as well as factor loadings (Gnambs & Staufenbiel, 2016). Becker and Wu (2007), for example, proposed a method for using regression coefficients as the effect sizes in a multivariate meta-analysis. This approach takes the dependency of the regression coefficients effectively into account. Also, study-level moderators can be included in the multivariate meta-analysis to explain differences in regression coefficients across studies. As pointed out by the authors, a limitation of

---

[1] One reviewer remarked that the use of the terms correlation- versus parameter-based MASEM is a poor description and imprecise use of the terminology because all models rely on "estimated" parameters, not the actual parameters themselves. For example, the correlation coefficients in Stage 1 of correlation-based MASEM are also parameter estimates. We acknowledge the potential confusions of these terms. However, because it is necessary to differentiate these two types of meta-analysis in MASEM and these terms have been used before, we continue using these terms in this study. "Correlation-based" indicates that the bivariate correlation coefficients are meta-analyzed, while "parameter-based" indicates that parameters of the structural equation model (e.g. path coefficients) are meta-analyzed.

Table 1

Overview of Possibilities of Different Random-Effects MASEM Methods

MASEM methods	Heterogeneity (correlations or SEM parameters)	Moderators	Missing correlations
Univariate-r approach (Viswesvaran & Ones, 1995)	Correlations	Multiple-group analysis with categorical moderators	Yes
GLS (Becker, 1992)	Correlations	Multiple-group analysis with categorical moderators	Yes
TSSEM (Cheung, 2014a)	Correlations	Multiple-group analysis with categorical moderators	Yes
FIMASEM (Yu, Downes, Carter, & O'Boyle, 2016)	SEM parameters	Unclear how moderators are handled	Yes
Parameter-based MASEM (Cheung & Cheung, 2016)	SEM parameters	Categorical and continuous moderators on the SEM parameters	No
Bayesian MASEM (Ke, Zhang, & Tong, 2018)	SEM parameters	Unclear how moderators are handled (but theoretically possible)	Yes
One-stage MASEM	Correlations	Categorical and continuous moderators on the SEM parameters	Yes

Note. MASEM = meta-analytic structural equation modeling; SEM = structural equation modeling; GLS = generalized least squares; TSSEM = two-stage SEM; FIMASEM = Full Information MASEM. Heterogeneity: Does the method provide estimates of the heterogeneity in the correlation coefficients or the parameters in the SEM model? Moderators: Does the method allow for categorical or continuous moderators explaining (part of) the heterogeneity of SEM parameters? Missing correlations = Can the method handle primary studies with incomplete correlation matrices?

this approach is that the sampling covariance matrix among the regression coefficients is often not reported. Researchers may, therefore, have to assume that the regression coefficients are uncorrelated, or plug in an arbitrary correlation. Moreover, the estimated regression coefficients may not be comparable across studies unless the same set of predictors are used in all studies, or the predictors are uncorrelated. Considering multiple regression as one of the simplest MASEM models, one could say that our new approach extends the earlier work of Becker and Wu (2007) to a broader range of structural equation models, such as factor models.

Parameter-based MASEM can provide estimates of the heterogeneity of parameters in the SEM, and allows for the inclusion of continuous moderator variables to explain this heterogeneity. However, with the exception of Bayesian MASEM which is discussed in the next paragraph, each primary study has to provide either all SEM-parameters of interest and their sampling covariance matrix, or complete data on the correlation matrices in order to fit the hypothesized model to the data of each study (see row 5 in Table 1). In practice, complete data is hardly ever available for any study (Sheng et al., 2016). Another problem with parameter-based MASEM is that the hypothesized model may not fit well in the individual studies, leading to biased parameter estimates being included in the meta-analysis.

Recently, Ke, Zhang, and Tong (2018) presented a promising parameter-based MASEM model in the Bayesian framework. Their method overcomes the limitations related to missing data from other parameter-based MASEM methods, and allows the evaluation of model fit using Bayesian fit measures, specifically the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) and posterior predictive checking (Gelman, Meng, & Stern, 1996). Ke et al. (2018) focused on the estimation of the heterogeneity in the SEM-parameters without explaining this between-studies variance. In principle, their model can be extended in order to include moderating variables, but the authors left this as a suggestion for future research. Therefore, applied researchers are not able to evaluate moderating effects on SEM parameters using this approach yet (see row 6 in Table 1). Also, the model specification of Bayesian MASEM might be challenging for the applied user, even with the example scripts in OpenBUGS that the authors supplied.

One-Stage Fixed-Effects MASEM

One-stage methods fit the meta-analytic structural equation model at the data from the primary studies directly. Oort and Jak (2016) presented a one-stage approach using maximum likelihood estimation for the fixed-effects model. Although this method is a fixed-effects method, we explain the model because of its similarities with the one-stage MASEM method that we propose in the next paragraph. In the Oort and Jak (2016) procedure, the structural equation model is fitted to the observed correlation matrices in a multigroup model, where each study is a group, and the parameter estimates of the structural model are constrained to be equal across groups/studies. Missing variables in the individual studies are filtered out when needed. This method stays within Stage 1 of TSSEM but restricts the pooled correlation matrix to adhere to the structure of the hypothesized structural equation model. A measure of fit is obtained by fitting a saturated model in addition to the hypothesized model and comparing the model loglikelihoods. The big disadvantage of the Oort and Jak (2016) procedure is that it only allows estimation of fixed-effect models, so it does not account for study-level heterogeneity, nor does it allow the inclusion of continuous moderator variables to explain the heterogeneity. The new method proposed in the next paragraph could be viewed as the random-effects extension of Oort and Jak (2016).

The New Method: One-Stage MASEM

The one-stage MASEM allows researchers to incorporate moderator variables into their MASEM, in which all parameters in the SEM can be modeled by the moderator variable, and that does not require complete data for the primary studies included in the meta-analysis (see row 6 of Table 1). A conceptual overview of one-stage MASEM is provided in Figure 3. Apart from seeing the method as the random-effect extension of the Oort and Jak (2016) procedure, one-stage MASEM can also be viewed as an extension of the random-effects meta-analysis of correlation matrices in Stage 1 of TSSEM (Cheung, 2014a). By treating studies in a meta-analysis as subjects in SEM, the vector of correlation coefficients and their variance component of the heterogeneity in a meta-analysis are modeled as the mean and covariance structures in SEM. Cheung only considered the unstructured (saturated)
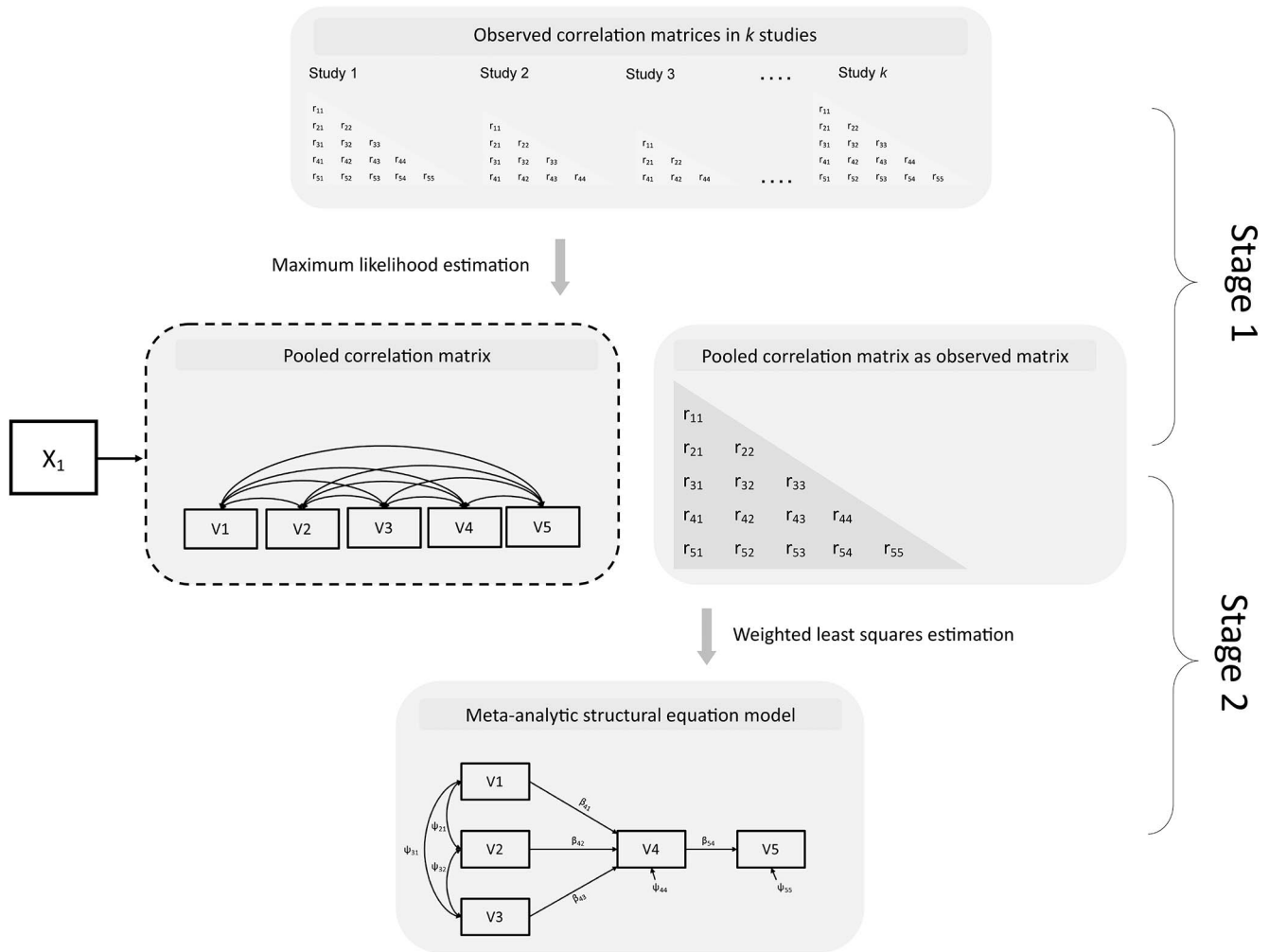
Figure 2. Conceptual representation of the TSSEM approach. The arrow pointing to the dashed box of the pooled correlation matrix indicates that in TSSEM, the pooled bivariate correlations can be specified to be a function of the moderator variable X.

model on the mean structure of the correlation coefficients. One-stage MASEM is an extension of this approach by imposing a model implied correlation structure on the mean structure.

Models Without Any Moderator

In one-stage MASEM, we consider the correlation coefficients as the "variables" and the studies as "subjects" in the data set.[2] Suppose the analysis involves 5×5 correlation matrices with k studies; we create a data matrix of k "subjects" (studies) and 10 "variables" (correlation coefficients). When there are incomplete data, they are handled by the use of the full information maximum likelihood (FIML) in the analysis.

Let us provide a brief overview of how missing correlations are handled with FIML in both TSSEM and one-stage MASEM. Rubin (1987) provided a framework to classify missing mechanisms in primary data. Pigott (2009) applied Rubin's (1987) definitions to meta-analysis. If the values of the missing correlations are not related to the missing values or other observed

variables in the analysis, the data are missing completely at random (MCAR). For example, a researcher does not include a correlation in his or her design because he or she is not interested in that correlation. The data are missing at random

---

[2] Mehta and Neale (2005) introduced the analogy of people as variables to illustrate how single level structural equation models could be used to analyze multilevel structural equation models. "Practically, this involves switching subscripts of a univariate MLM (people and clusters) with those of a CFA model (variables and people); that is, individual scores are conceptualized as separate variables, and the unit of analysis for the SEM model is now the cluster" (Mehta & Neale, 2005, p. 263). In a series of articles, Cheung (2008, 2013, 2014b) extended this framework to the univariate, multivariate, and three-level meta-analyses. Specifically, studies (or independent clusters) in a meta-analysis are conceptualized as subjects in a structural equation model. The known within-study sampling variance-covariance matrix $\mathbf{V}_i$ among the effect sizes is properly handled by imposing it as fixed values via definition variables (see Cheung, 2013 for the details).
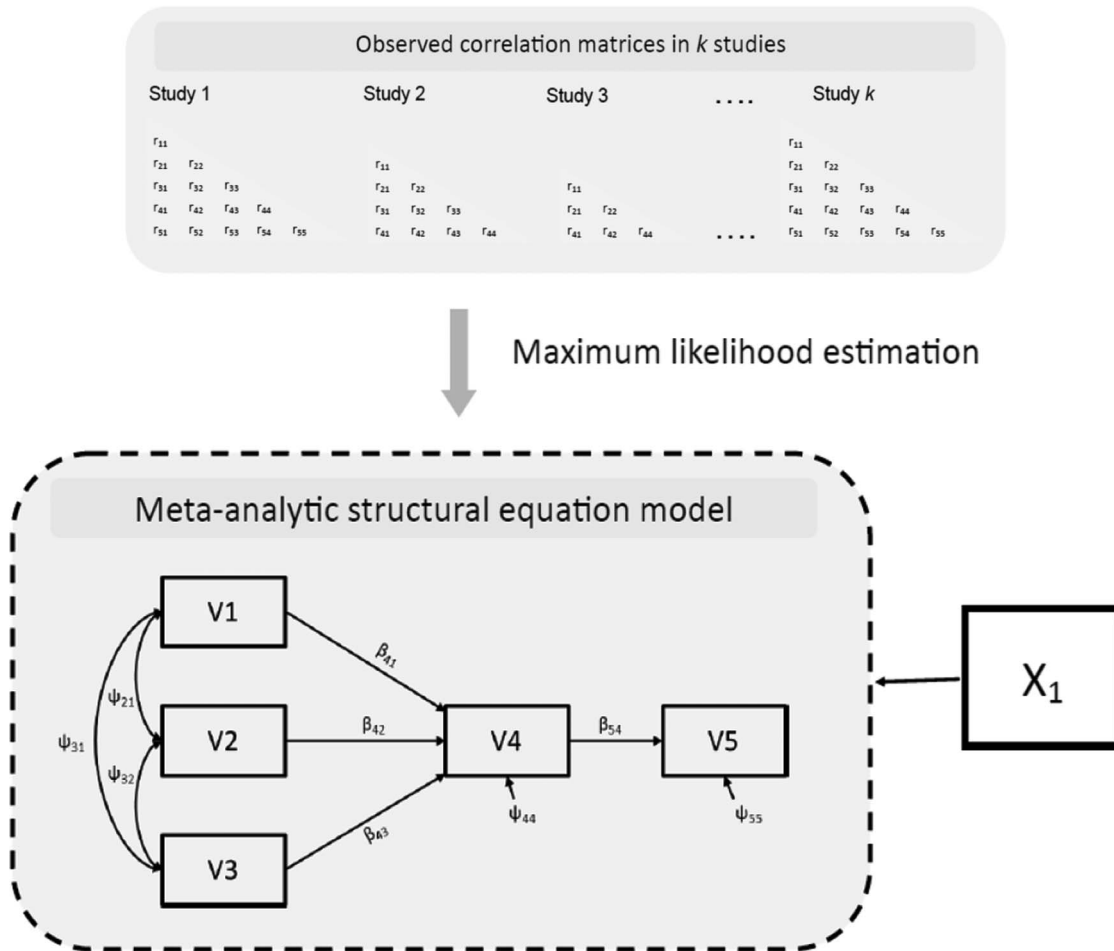
Figure 3. Conceptual representation of one-stage MASEM. The arrow pointing to the dashed box of the structural equation model indicates that all parameters of the hypothesized model can be specified to be a function of moderator variable X.

(MAR) if the missingness can be fully accounted for by the other variables in the analysis. For example, a researcher does not report a correlation because other correlations are very high (or low). On the other hand, the data are missing not at random (MNAR) if the missingness is related to the values of the missing correlations even after controlling for other variables in the analysis. For example, a researcher does not report a correlation because that specific correlation is very high (or low). As the incomplete correlations in MASEM are most likely created by design (some researchers were interested in including this set of variables whereas others were not), it may often be reasonable to assume that the missing correlations are MCAR (Yuan & Kano, 2018). On the other hand, if researchers do not report some of the correlations because these correlations are small or nonsignificant, the assumption of MCAR may not be valid.

Both multiple imputation (MI) and FIML are unbiased and efficient in handling missing data with either MCAR or MAR (e.g., Enders, 2010; Graham, 2009; Schafer & Graham, 2002). When the missingness mechanism is MNAR, FIML (and MI) are not unbiased. However, the bias of FIML is generally

smaller than those in other ad hoc methods such as the listwise or pairwise deletion (Jamshidian & Bentler, 1999; Muthén, Kaplan, & Hollis, 1987). In fact, many missing data experts still prefer to apply methods assuming MAR even the data are MNAR in primary research. Enders (2010, p. 344), for example, stated "a well-executed MAR analysis may be preferable to an MNAR analysis, even if there is a reason to believe that missingness is systematically related to the outcome variable." All current MASEM techniques such as Becker (1992), Cheung and Chan (2004; Cheung, 2014a), and Viswesvaran and Ones (1995)[3] assume that the missing correlations are either MCAR or MAR. Therefore, it may be reasonable to apply FIML in

---

[3] Viswesvaran and Ones (1995) state that the population matrix should be filled with the best possible estimates. As one reviewer remarked, in case of missing correlation coefficients in primary studies due to selective reporting (publication bias), leading to MNAR data, one could in principle plug in estimates adjusted for publication bias. It is, however, not evident how such an estimate should be obtained.

MASEM as well, although researchers should evaluate the missing data assumptions in their specific MASEM analyses.

Let us illustrate the key idea of FIML with the model in Figure 1a. There are five variables in the structural equation model ($5 \times 4/2 = 10$ means in $\boldsymbol{\mu}(\boldsymbol{\theta})_i$ and a 10×10 covariance matrix in $\boldsymbol{\Sigma}(\boldsymbol{\theta})_i$) when there is no missing data. Please note that there is a subscript i in the model implied means and covariance matrix meaning that their dimensions may vary across studies. Suppose that there is a data point (correlation coefficient) missing in the second study. A filter matrix is created to remove the relevant dimensions with that missing value. Therefore, there will be only nine elements in $\boldsymbol{\mu}(\boldsymbol{\theta})_2$ and the dimension of $\boldsymbol{\Sigma}(\boldsymbol{\theta})_2$ is 9×9. The log-likelihood of the model is calculated by summing the individual log-likelihoods with possibly different dimensions. The model fit, parameter estimates, and their standard errors are computed by maximizing the log-likelihood of the model. Readers may refer to Enders (2010) for a comprehensive treatment on this subject.

One-stage MASEM fits the SEM by restricting the pooled correlations in the (multivariate) random-effects model. In its simplest form, the random-effects model decomposes the vector $\mathbf{r}_i$ of observed correlation coefficients for a study i in three parts:

$$\mathbf{r}_i = \boldsymbol{\rho}_R + \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\boldsymbol{\rho}_R$ indicates the mean vector of the correlation coefficients, $\mathbf{u}_i$ is a vector of deviations of study i's population correlation coefficients from $\boldsymbol{\rho}_R$, and $\boldsymbol{\varepsilon}_i$ is a vector with the sampling error of study i. $\mathrm{Cov}(\mathbf{u}_i) = \mathbf{T}^2$ denotes the between-studies covariance matrix that has to be estimated, and $\mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \mathbf{V}_i$ denotes the sampling covariance matrix of the correlation coefficients, which is usually treated as known in a meta-analysis (see Olkin and Siotani (1976) for the formula in computing $\mathbf{V}_i$).[4]

The structural model, for example, a path model, is nested under the model in Equation 1 and is obtained by restricting $\boldsymbol{\rho}_R$:

$$\boldsymbol{\rho}_R = \mathrm{vechs}(\mathbf{F}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{S}(\mathbf{I} - \mathbf{A})^{-1T}\mathbf{F}^T), \tag{2}$$

where using the RAM-formulation (McArdle & McDonald, 1984), $\mathbf{I}$ is an identity matrix, $\mathbf{F}$ is a selection matrix with 1's for observed variables and 0's for latent variables, $\mathbf{A}$ is a square matrix with asymmetric paths such as regression coefficients and factor loadings, $\mathbf{S}$ is a symmetrical matrix with variances and covariances, and vechs() vectorizes the lower diagonal of its argument. If there are 5×5 correlation matrices in the analysis, $\boldsymbol{\rho}_R$ is a 10×1 column vector whereas $\mathbf{T}^2$ and $\mathbf{V}_i$ are two 10×10 matrices. To illustrate the RAM-model formulation, for the path model in Figure 1a, the matrices $\mathbf{A}$, $\mathbf{S}$, and $\mathbf{F}$ would look as follows when the variables are arranged as social norms, attitudes, perceived control, smoking intention, and smoking:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{4,1} & a_{4,2} & a_{4,3} & 0 & 0 \\ 0 & 0 & a_{5,3} & a_{5,4} & 0 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} 1 & & & & \\ s_{2,1} & 1 & & & \\ s_{3,1} & s_{3,2} & 1 & & \\ 0 & 0 & 0 & s_{4,4} & \\ 0 & 0 & 0 & 0 & s_{5,5} \end{bmatrix}, \text{ and}$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In general, for models with p observed variables, $\boldsymbol{\rho}_R$ will be a $p \times (p - 1)/2$ dimensional column vector, and $\mathbf{A}$, $\mathbf{S}$, and $\mathbf{F}$ will be of dimensions $p \times p$. For models with p observed variables and q latent variables, matrices $\mathbf{A}$ and $\mathbf{S}$ will be of dimensions $(p + q) \times (p + q)$, matrix $\mathbf{F}$ will be $p \times (p + q)$ and filters out the latent variables, so that $\boldsymbol{\rho}_R$ will be a $p \times (p - 1)/2$ dimensional column vector. For more explanation of the RAM-formulation, see McArdle (2005) and Loehlin and Beaujean (2016).

Because we are using correlation matrices as inputs, the variances of the exogenous variables are fixed at 1. Moreover, the diagonal elements of the model implied correlation matrix in Equation 2 should always be equal to one during estimation (Cudeck, 1989). Either imposing constraints on the parameters or reparameterizing the model is required to ensure that all the diagonal elements are equal to 1 (see Section 7.3.2.1 of Cheung [2015a]).

All model parameters are estimated with FIML as implemented in the metaSEM package (Cheung, 2015b) which uses the OpenMx package (Neale et al., 2016) as the backend in the R statistical platform (R Core Team, 2019). A test statistic of the hypothesized model can be obtained by performing a likelihood ratio test with the saturated model from Equation 1, similar to Oort and Jak (2016). A detailed graphical display of one-stage MASEM is provided in Figure 4. In order to facilitate the comparison with TSSEM, we also provide a detailed graphical overview of TSSEM in Figure 5.

Moderation of the Regression Coefficients in the A Matrix

Bauer (2017) presented a method that might model any SEM-parameter as a function of a moderator variable in the context of measurement invariance.[5] Conceptually, the inclusion of moderators to explain variation in the SEM-parameters across studies with one-stage MASEM resembles the approach of Bauer (2017). Moderator variables can be included in one-stage MASEM to explain (part of the) heterogeneity across studies, by regressing model parameters on moderator variables. In principle, all model parameters in the $\mathbf{A}$ and $\mathbf{S}$ matrices can be regressed on the moderators. We expect that in practice, researchers are most often interested in moderation of $\mathbf{A}$. Regressing the direct effects in $\mathbf{A}$ on moderator variable X would be modeled as:

---

[4] Olkin and Siotani (1976) provided an analytic formula to estimate $\mathbf{V}_i$. The metaSEM package, which is used in this article, uses an SEM approach to compute $\mathbf{V}_i$ (Cheung & Chan, 2004). Because both approaches use maximum likelihood estimation with the assumption of multivariate normal on the data, the results are identical.

[5] In testing the measurement invariance across groups, researchers usually choose between a multiple-group SEM approach and a multiple-indicator multiple-cause (MIMIC) approach. These two approaches have their own pros and cons. For example, the multiple-group approach can be used to model any between-group differences on the intercepts of the variables, factor loadings, and latent variance-covariance matrices. However, the multiple-group approach can only handle categorical moderators such as gender and ethnicity. On the other hand, the MIMIC model allows both categorical and continuous moderators such as age. The limitation of the standard MIMIC model is that only the intercepts of the variables can be modeled. Bauer (2017) resolved this dilemma by presenting a new approach that can use both categorical and continuous moderators to model all the parameters in factor models.
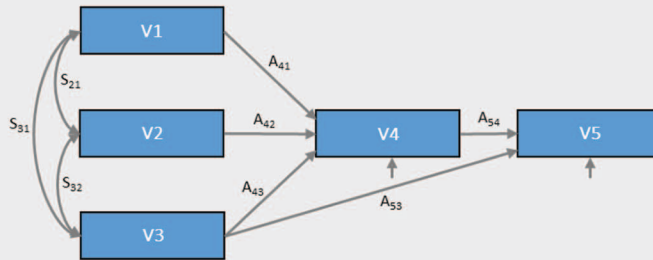
One-Stage MASEM

Restricting the pooled correlations to be a function of SEM model parameters

Data:

	r_{21}	r_{31}	r_{41}	r_{51}	r_{32}	r_{42}	r_{52}	....	r_{54}
Study 1	x	x	x	x					x
Study 2		x	x	x				x	x
...	x	x		x			x		
Study k	x	x	x	x	x	x	x	x	x

Hypothesized model:

V1, V2, V3, V4, V5 with paths S_{21}, S_{31}, S_{32}, A_{41}, A_{42}, A_{43}, A_{53}, A_{54}

$$P(\theta) = F (I - A)^{-1} S (I - A)^{-1T} F^T$$

$$P(\theta) = \begin{vmatrix} 1 \\ \rho_{21}(\theta) & 1 \\ \rho_{31}(\theta) & \rho_{32}(\theta) & 1 \\ \rho_{41}(\theta) & \rho_{42}(\theta) & \rho_{43}(\theta) & 1 \\ \rho_{51}(\theta) & \rho_{52}(\theta) & \rho_{53}(\theta) & \rho_{53}(\theta) & 1 \end{vmatrix}$$

Model fitting:

$$\mu(P(\theta)) = \text{vechs}(P(\theta))$$
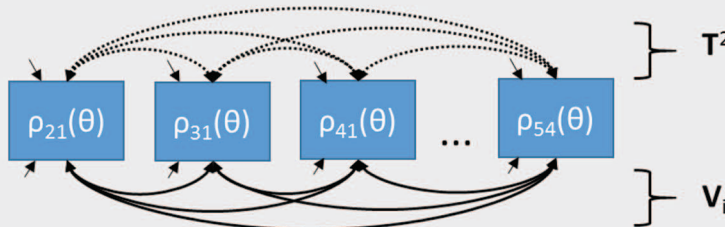$$\Sigma(P(\theta)) = V_i + T^2$$

$\rho_{21}(\theta)$ $\rho_{31}(\theta)$ $\rho_{41}(\theta)$ ... $\rho_{54}(\theta)$ ⎫ T^2

⎫ V_i

Figure 4. Detailed graphical representation of one-stage MASEM. See the online article for the color version of this figure.

Two-Stage SEM

Stage 1: Pooling correlations across studies

Data:

	r_{21}	r_{31}	r_{41}	r_{51}	r_{32}	r_{42}	r_{52}	....	r_{54}
Study 1	x	x	x	x					x
Study 2		x	x	x			x	x
...	x	x		x			x	
Study k	x	x	x	x	x	x	x	x	x

Maximum likelihood estimation

Pooled correlations and variance components:

$$\mu(P) = (\rho_{21}, \rho_{31}, \rho_{41}, \rho_{51}, \rho_{32}, \rho_{42}, \rho_{52}, ..., \rho_{54})$$
$$\Sigma(P) = V_i + T^2$$

Acov: Asymptotic covariance matrix of pooled correlations

Stage 2: Fitting the SEM to the pooled correlations

Pooled correlations as observed matrix:

r_{11}
r_{21} r_{22}
r_{31} r_{32} r_{33}
r_{41} r_{42} r_{43} r_{44}
r_{51} r_{52} r_{53} r_{54} r_{55}

Weighted Least Squares estimation (using $ACOV^{-1}$)
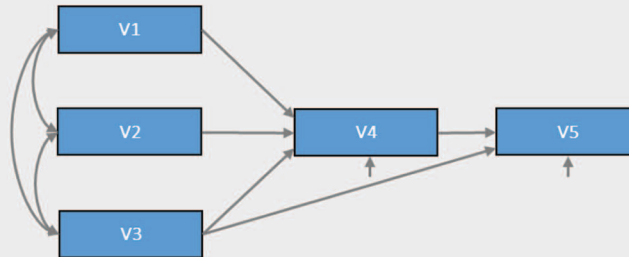
Hypothesized model:

V1

V2 V4 V5

V3

Figure 5. Detailed graphical representation of TSSEM. See the online article for the color version of this figure.

$$\mathbf{A}_i = \mathbf{A}_0 + \mathbf{A}_1 \circ \mathbf{X}_i, \qquad (3)$$

where $\mathbf{A}_i$ is the matrix with the *i*th study-specific asymmetric paths, $\mathbf{X}_i$ is a matrix with the *i*th study-specific values of the moderator variable X, $\mathbf{A}_1$ is a matrix with the regression effects of the moderator variable(s), $\mathbf{A}_0$ represents a matrix with intercept values for the asymmetric paths when $\mathbf{X}_i = 0$, and the $\circ$-operator denotes the Hadamard or element-wise product. Taking the model from Figure 1a as an example, the model matrices would look as follows:

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{0,4,1} & a_{0,4,2} & a_{0,4,3} & 0 & 0 \\ 0 & 0 & a_{0,5,3} & a_{0,5,4} & 0 \end{bmatrix},$$

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{1,4,1} & a_{1,4,2} & a_{1,4,3} & 0 & 0 \\ 0 & 0 & a_{1,5,3} & a_{1,5,4} & 0 \end{bmatrix}, \text{ and}$$

$$\mathbf{X}_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_i & x_i & x_i & 0 & 0 \\ 0 & 0 & x_i & x_i & 0 \end{bmatrix}.$$

Note that in this model, the value of the moderator can be different for each coefficient in theory, allowing the inclusion for effect-size level moderators. However, we expect that in practice, the moderator will be study-level so that all values in $\mathbf{X}_i$ are equal within the same study. Similar to creating an interaction term for continuous variables, the results will be more stable if we center or standardize $\mathbf{X}_i$ before the analysis.

The complete model (obtained by plugging in Equation 3 in Equation 2 in Equation 1) then becomes:

$$\mathbf{r}_i = \text{vechs}(\mathbf{F}(\mathbf{I} - (\mathbf{A}_0 + \mathbf{A}_1 \circ \mathbf{X}_i))^{-1}\mathbf{S}(\mathbf{I} - (\mathbf{A}_0 + \mathbf{A}_1 \circ \mathbf{X}_i))^{-1\mathrm{T}}\mathbf{F}^{\mathrm{T}})$$
$$+ \mathbf{u}_i + \boldsymbol{\varepsilon}_i. \qquad (4)$$

$\text{Cov}(\mathbf{u}_i) = \mathbf{T}^2$ in this model represents the *residual* between-study covariance matrix. Therefore, the model implied mean and covariance structure of the *i*th study are:

$$\boldsymbol{\mu}(\boldsymbol{\theta})_i = \text{vechs}(\mathbf{F}(\mathbf{I} - (\mathbf{A}_0 + \mathbf{A}_1 \circ \mathbf{X}_i))^{-1}$$
$$\times \mathbf{S}(\mathbf{I} - (\mathbf{A}_0 + \mathbf{A}_1 \circ \mathbf{X}_i))^{-1\mathrm{T}}\mathbf{F}^{\mathrm{T}}), \text{ and} \qquad (5)$$
$$\boldsymbol{\Sigma}(\boldsymbol{\theta})_i = \mathbf{T}^2 + \mathbf{V}_i.$$

A comparison with the $\mathbf{T}^2$ of the model without moderators will provide insight in how much of the between-study variance in the observed correlation coefficients is explained by the moderators. Suppose $\mathbf{T}_0^2$ and $\mathbf{T}_1^2$ are the heterogeneity (co)variance matrices from a model with and without the moderators. We may calculate an $R_{1,1}^2$ on the first correlation coefficient by using the following formula, where the indices in the bracket indicate the elements in the matrices:

$$(\mathbf{T}_1^2[1, 1] - \mathbf{T}_0^2[1, 1])/\mathbf{T}_0^2[1, 1]. \qquad (6)$$

Negative values of R^2 are usually truncated to zero. This definition resembles the standard explained variance used in mixed-effects meta-analysis (e.g., Borenstein et al., 2009).

Moderation of the Covariances in the S Matrix

Although most research questions will focus on the moderation of direct effects in path models, researchers may still test moderating effect on the covariances in $\mathbf{S}$ by regressing $\mathbf{S}$ on moderator variable X as:

$$\mathbf{S}_i = \mathbf{S}_0 + \mathbf{S}_1 \circ \mathbf{X}_i, \qquad (7)$$

where $\mathbf{S}_i$ is the matrix with the *i*th study-specific symmetric paths, $\mathbf{X}_i$ is a matrix with the *i*th study-specific values of the moderator variable X, $\mathbf{S}_1$ is a matrix with the regression effects of the moderator variable(s), $\mathbf{S}_0$ represents a matrix with intercept values for the symmetric paths, and the $\circ$-operator denotes the Hadamard or element-wise product.

Because correlation structures are fitted in MASEM, the elements on the diagonal of the $\mathbf{S}$ are no free parameters. Thus, only the off-diagonal elements in $\mathbf{S}$ can be moderated by X. Taking the model from Figure 1a as an example would lead to the following model matrices:

$$\mathbf{S}_0 = \begin{bmatrix} 1 & & & & \\ s_{0,2,1} & 1 & & & \\ s_{0,3,1} & s_{0,3,2} & 1 & & \\ 0 & 0 & 0 & s_{0,4,4} & \\ 0 & 0 & 0 & 0 & s_{0,5,5} \end{bmatrix},$$

$$\mathbf{S}_1 = \begin{bmatrix} 0 & & & & \\ s_{1,2,1} & 0 & & & \\ s_{1,3,1} & s_{1,3,2} & 0 & & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ and}$$

$$\mathbf{X}_i = \begin{bmatrix} 0 & & & & \\ x_i & 0 & & & \\ x_i & x_i & 0 & & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Similar to a model where the elements in $\mathbf{A}$ are moderated, the $\mathbf{T}^2$ of the models with and without moderators can be used to estimate how much of the between-studies variance in the observed correlation coefficients is explained by the moderators.

Illustrations

We will present two illustrations of one-stage MASEM with moderators using two real data sets. The first illustration evaluates a path model, and the second illustration evaluates a factor model. A detailed and annotated overview of scripts and outputs for all analyses are included as online supplementary materials. The latest version of the R code is also available in Github (https://github .com/mikewlcheung/code-in-articles).

Illustration 1: Work–Family Conflict and Strain

Nohe, Meier, Sonntag, and Michel (2015) meta-analyzed data from 32 independent samples from 30 panel studies on the relationship between work–family conflict and strain, in order to evaluate whether work–family conflict predicts strain, strain predicts work–family conflict, or whether they are reciprocally related. The authors originally analyzed their data using univariate meta-analysis to obtain the average correlation coefficients and

subsequently treated the average correlation matrix as input in a standard SEM-analysis (the univariate-r method). The theoretical model that the researchers were interested in is the two-wave cross-lagged panel model as depicted in Figure 6. This model includes two autoregressive effects, strain to strain and work to work (β_{S2S} and β_{W2W}), and two cross-lagged effects, work to strain and strain to work (β_{W2S} and β_{S2W}) between strain and work-family conflict.

The authors evaluated the moderation effect of time lag on the bivariate correlation coefficients, using a series of univariate metaregressions. However, the actual interest was not in moderation of the bivariate correlations, but in moderation of the regression coefficients between waves (see Figure 6). In this illustration, we will reanalyze their data using one-stage MASEM with a time lag as a moderator of the regression coefficients. For comparison, we will analyze the same data with TSSEM, which should give similar results for the model without moderators.

Data. The dataset consists of 32 four by four correlation matrices with the bivariate correlations between work–family conflict and strain; both measured at two-time points (these data are obtained from Table A1 in the article by Nohe, Meier, Sonntag, & Michel, 2015). The associated sample sizes ranged from 66 to 2,235, with an average of 403.3 and a median of 247.5. Studies differed in the time lag between the two waves. The average time lag was 14.12 months, and the median time lag was 12 months, with a range from around 1 week to 72 months.

Analysis. We used the R-package metaSEM (Cheung, 2015b), which includes dedicated functions to apply TSSEM as well as one-stage MASEM. We fitted the model from Figure 6 to the observed correlation matrices in order to evaluate the regression coefficients. Note that because the model is saturated, the model fit is perfect by definition. Next, we included time lag as a potential moderator of the regression coefficients. It was expected that studies with a longer time interval between the two waves would find smaller effects across time. With one-stage MASEM, the variable "lag" is included as a continuous moderator. For comparison, we also analyzed the effect of lag using subgroup

TSSEM (Jak & Cheung, 2018b). We categorized the variable by creating three subgroups of studies based on the time lag: 1–6 months, 7–12 months, and more than 13 months. This categorization corresponds to the subgroups that the authors analyzed in the original meta-analysis. Next, we fitted the path model to all three groups simultaneously and used a chi-square difference test between a model with and without the regression coefficients constrained to be equal across subgroups. We used a significance level of $\alpha = .05$ in this study.

Results. In the following section, we first present the results of the overall analysis and then the results of the moderation analysis.

Overall analysis. One-stage MASEM and TSSEM provide the same parameter estimates and standard errors for the overall model (see the first two columns of Table 2). The two cross-lagged effects are both positive, statistically significant, and of similar small size (0.080 and 0.086). These results support the idea that the effects of strain and work–family conflict are reciprocal (Nohe et al., 2015). The autoregressive effects are both positive, of medium size, and also statistically significant (0.586 for strain and 0.572 for work–family conflict).

Moderation analysis. We added the variable "lag," which was standardized to improve numerical stability, as a moderator explaining variance in the four regression coefficients across studies. The omnibus test of the moderation effects was statistically significant, $\chi^2(4) = 23.522$, $p < .05$. All four moderation effects were in the expected direction (negative), although only the autoregressive effect of work–family conflict was statistically significant (see the third column of Table 2). The intercept of this regression coefficient is 0.573, which corresponds to a study with the mean time lag (14.12 months). For each standard deviation of months increase in a time lag, the estimated autoregressive effects are expected to decrease with 0.062 points. The calculation of R^2 with the between-studies variance of the model with and without the moderator shows that 40.7% of the variance in the correlation coefficient between work-family conflict at T1 and T2 could be explained by lag.
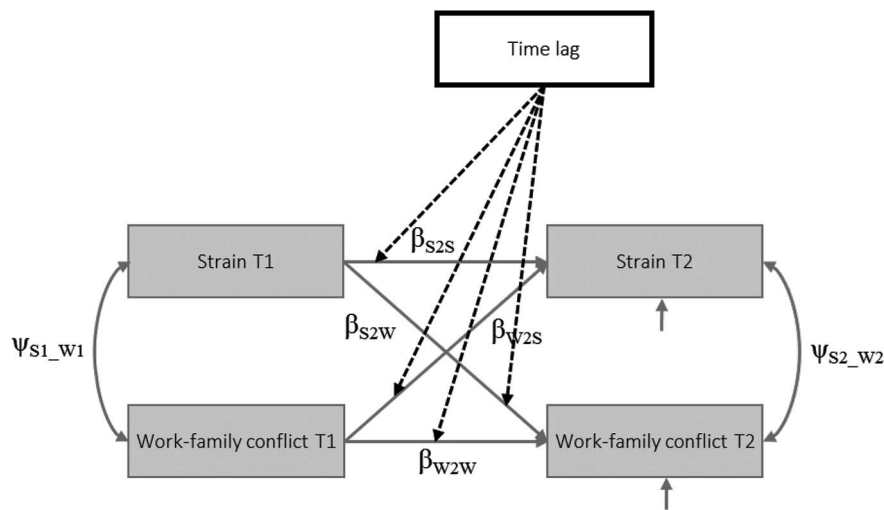

Figure 6. Hypothesized panel model from Nohe et al. (2015), with time lag as a moderator of the regression coefficients.

Table 2

Estimated Path Coefficients With Standard Errors, Variance Estimates, and Explained Variance (R^2) of the Analyses With TSSEM and One-Stage MASEM on the Dataset of Nohe et al. (2015)

Item	TSSEM Estimate (*SE*)	One-stage MASEM Estimate (*SE*)	One-stage MASEM with moderator Estimate (*SE*)	R^2
β_{S2S}	.586 (.021)	.586 (.021)	.586 (.020)	
β_{W2W}	.572 (.022)	.572 (.022)	.573 (.018)	
β_{W2S}	.080 (.025)	.080 (.025)	.080 (.024)	
β_{S2W}	.086 (.025)	.086 (.025)	.085 (.024)	
Ψ_{W1_S1}	.380 (.023)	.380 (.023)	.381 (.023)	
Ψ_{W2_S2}	.169 (.025)	.169 (.025)	.167 (.025)	
Lag $\rightarrow \beta_{S2S}$			−.028 (.020) ns	
Lag $\rightarrow \beta_{W2W}$			−.062 (.019)	
Lag $\rightarrow \beta_{W2S}$			−.026 (.021) ns	
Lag $\rightarrow \beta_{S2W}$			−.002 (.021) ns	
$\tau^2_{W1\_S1}$	.013	.013	.014	.000
$\tau^2_{W1\_W2}$	.009	.009	.005	.407
$\tau^2_{W1\_S2}$	.007	.007	.006	.097
$\tau^2_{S1\_W2}$	.007	.007	.006	.083
$\tau^2_{S1\_S2}$	.007	.007	.006	.139
$\tau^2_{W2\_S2}$	.012	.012	.012	.000

Note. TSSEM = two-stage SEM; MASEM = Meta-analytic structural equation modeling. ns indicates that the parameter estimate not statistically significant at α = .05. The between-studies variance (τ^2) of the correlation coefficient in TSSEM is estimated from the Stage 1 analysis. R^2 denotes the proportion of between-studies variance (τ^2) in the correlation coefficient that is explained by the regression of β on Lag. It is calculated as $1 - (\tau^2_1/\tau^2_0)$, where τ^2_1 and τ^2_0 represent the τ^2 of the model with and without the moderator respectively. Negative values are truncated to zero. ns indicates that the parameter estimate is not statistically significant at α = .05.

As a comparison, we also tested moderation using subgroup TSSEM. Constraining the four regression coefficients to be equal across the three subgroups of studies did not significantly deteriorate model fit, $\chi^2(8) = 14.738$, $p = .064$, so for this analysis, the hypothesis of equal regression coefficients across lags would not be rejected. Testing the equality of only the regression coefficient β_{W2W}, which was found to be significantly moderated with the one-stage MASEM, also did not result in rejection of the null hypothesis of equality of the effect across subgroups, $\chi^2(2) = 2.527$, $p = .283$. To demonstrate that the one-stage MASEM works with incomplete data as well, we randomly dropped one of the four variables in each study. This means that three out of six correlations (half of the data) are dropped in each study. As can be seen in the online supplementary materials, the results are comparable to those with complete data. As expected, the standard errors of the findings with incomplete data tend to be larger because of the smaller numbers of studies.

Conclusion. This illustration showed that one-stage MASEM and TSSEM lead to the same parameter estimates in the models without moderators. When analyzing the effect of the moderator, we found different results. In TSSEM we needed to categorize the continuous moderator in order to create subgroups. As a consequence, the power to detect moderation effects decreased. Moreover, it is rather difficult to decide how to categorize the variable, especially if the results change based on different categorization options. In this example, we followed the procedure of the original authors, but we could also have performed a median split to create

two subgroups instead of three. Analyzing two subgroups instead of three would have led to a rejection of the null hypothesis of equal effects across the two subgroups $\chi^2(4) = 13.247$, $p = .010$. With one-stage MASEM on the other hand, we could include Lag as the original continuous variable that it was, without making arbitrary splits, and with the additional benefit that this procedure results in more statistical power to detect moderation effects.

Illustration 2: Factor Analysis of the Rosenberg Self-Esteem Scale

Gnambs, Scharl, and Schroeders (2018) examined the factor structure of the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965) using data from 113 independent samples from various countries. By applying fixed-effects TSSEM, they found that a bifactor model with one general self-esteem factor and two specific factors for positively and negatively worded items and fitted the combined data best. As a sensitivity analysis, a random-effects TSSEM was applied as well, leading to the same conclusions. The authors were interested in whether the factor loadings were moderated by the country's individualism score as defined by Minkov et al. (2017). This moderator hypothesis was tested using subgroup analysis by splitting the samples into two groups on the mean individualism score. It appeared that the factor loadings were generally larger in samples from more individualistic countries.

Gnambs et al. (2018) were forced to categorize the individualism variable based on some value (they used the mean in this case), while ideally the variable would have been evaluated using all available information. Moreover, the choice to use a fixed-effects approach is questionable, since homogeneity of correlation coefficients across samples is very unlikely. We reanalyze their data using one-stage MASEM with individualism as a continuous moderator. Figure 7 shows the hypothesized model, in which the dashed arrows from individualism to the factor loadings represent the moderation effects on the factor loadings.

Data. The RSES consists of 10 items. Half of the items are positively formulated (e.g., "On the whole, I am satisfied with myself") and the other half is negatively phrased (e.g., "At times, I think I am no good at all"). For 34 of the 113 samples, the bivariate correlations between the items were available, and individualism scores were reported. For the remaining samples, the authors calculated the bivariate correlations from the reported factor loadings (Gnambs & Staufenbiel, 2016). Because these reproduced correlations may be wrong if the specified factor structure is wrong, we will only use the 34 samples for which the bivariate correlations were directly available. The sample sizes of the individual samples ranged from 122 to 22,131, with a median sample size of 1,221. The individualism scores ranged from −126 (Philippines) to 180 (Netherlands) with a mean of 70, which was standardized in the analyses to improve numerical stability.

Analysis. Using one-stage MASEM, we fitted the hypothesized bifactor model to the observed correlation matrices. Then, we evaluated the fit of the factor model and interpreted the parameter estimates. The exact fit will typically be rejected in MASEM, due to the large total sample size. Therefore, as in standard SEM, it is common to evaluate the approximate fit of the hypothesized model. The approximate close fit is associated with RMSEA-values under .05, a satisfactory approximate fit with RMSEA-values under .08, and a bad approximate fit is associated
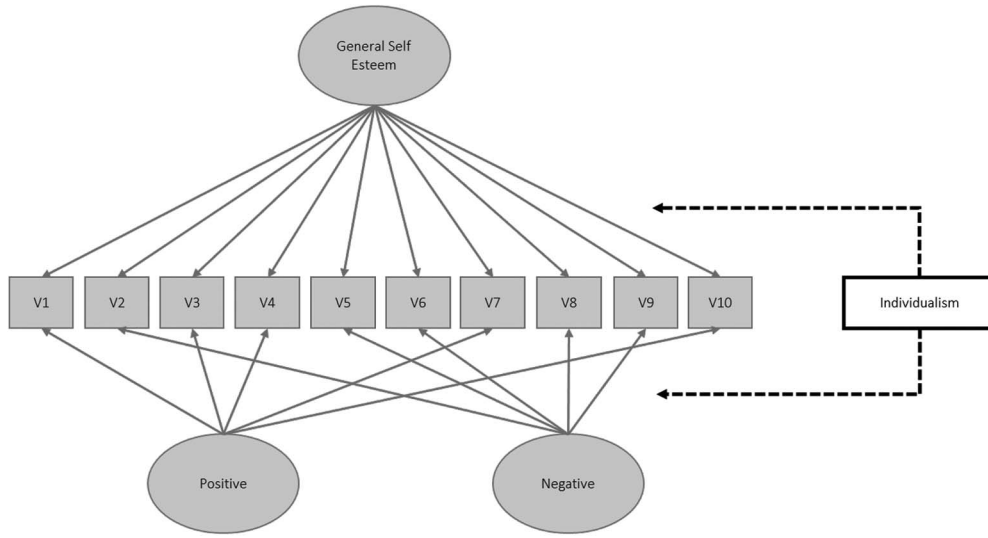
Figure 7. Bifactor model on self-esteem indicators from Gnambs et al. (2018), with individualism as a moderator of the factor loadings.

with RMSEA-values larger than .10 (MacCallum, Browne, & Sugawara, 1996). In addition to the RMSEA, we evaluated the standardized root mean squared residual (SRMR). SRMR-values under .08 are considered satisfactory (Hu & Bentler, 1999). Moderation of the factor loadings is evaluated by including individualism as a continuous moderator variable. Based on the results from the subgroup analyses in the original study, it was expected that factor loadings increase with larger individualism scores. For comparison with existing methods, we also analyzed the data using TSSEM. For the subgroup TSSEM, we split the samples based on the mean individualism score.

Results. In the following section, we first present the results of the overall analysis and then the results of the moderation analysis.

Overall analyses. The model fit of the bifactor model was good, with very similar to identical fit-statistics, parameter estimates, and standard errors for the two methods (see Table 3). Factor loadings for the general factor were all positive and significantly larger than 0, ranging from 0.39 to 0.80. Factor loadings for the specific factors were generally smaller, and two factor loadings for the positive factor were not statistically significant. These results are in line with the results the authors found in the original article.

Moderation analysis. We added individualism as a moderator of all factor loadings to the analysis. Overall, individualism had a statistically significant positive effect on the factor loadings, $\chi^2(20) = 270.883$, $p < .05$. The largest effects were found on the factor loadings of the general factor (see Table 4), for which six of

Table 3

Estimated Factor Loadings (With Standard Errors) and Model Fit of the Analyses With TSSEM and One-Stage MASEM on the Dataset of Gnambs et al. (2018)

	TSSEM		One-stage MASEM	
Item	General factor	Specific factor	General factor	Specific factor
1. On the whole, I am satisfied with myself (positive).	.753 (.014)	−.048 (.043) ns	.752 (.014)	−.046 (.044) ns
2. At times, I think I am no good at all (negative).	.536 (.013)	.588 (.024)	.536 (.013)	.588 (.025)
3. I feel that I have a number of good qualities (positive).	.596 (.017)	.531 (.070)	.595 (.018)	.533 (.071)
4. I am able to do things as well as other people (positive)	.522 (.013)	.306 (.038)	.522 (.013)	.307 (.038)
5. I feel I do not have much to be proud of (negative).	.527 (.014)	.326 (.022)	.528 (.016)	.323 (.023)
6. I certainly feel useless at times (negative).	.514 (.012)	.598 (.023)	.515 (.012)	.599 (.024)
7. I feel that I'm a person of worth, at least on an equal plane with others (positive).	.621 (.014)	.317 (.041)	.620 (.014)	.318 (.042)
8. I wish I could have more respect for myself (negative).	.385 (.018)	.400 (.030)	.385 (.019)	.399 (.032)
9. All in all, I am inclined to feel that I am a failure (negative).	.595 (.014)	.387 (.022)	.595 (.014)	.385 (.023)
10. I take a positive attitude toward myself (positive).	.802 (.014)	−.034 (.044) ns	.803 (.015)	−.034 (.045) ns
Fit statistics	$\chi^2(25) = 37.621$, $p = .050$		$\chi^2(25) = 35.938$, $p = .073$	
	RMSEA = .002		RMSEA = .002	
	95% CI [.000, .004]		95% CI [.000, .004]	
	SRMR = .016		SRMR = .016	

Note. TSSEM = two-stage SEM; MASEM = Meta-analytic structural equation modeling. ns indicates that the parameter estimate is not statistically significant at $\alpha = .05$.

Table 4

Estimated Moderating Effects (With Standard Errors) of Individualism of the Country on the Factor Loadings

Item	A₀ Intercept (General factor) Estimate (SE)	A₀ Intercept (Specific factor) Estimate (SE)	A₁ Slope (General factor) Estimate (SE)	A₁ Slope (Specific factor) Estimate (SE)
1. On the whole, I am satisfied with myself (positive)	.727 (.019)	.003 (.052) ns	.088 (.021)	−.162 (.034)
2. At times, I think I am no good at all (negative)	.551 (.019)	.569 (.028)	.022 (.020) ns	.007 (.030) ns
3. I feel that I have a number of good qualities (positive)	.569 (.021)	.505 (.044)	.073 (.016)	−.060 (.030)
4. I am able to do things as well as other people (positive)	.504 (.018)	.354 (.039)	.035 (.016)	−.042 (.026) ns
5. I feel I do not have much to be proud of (negative)	.546 (.021)	.300 (.031)	.001 (.021) ns	−.001 (.031) ns
6. I certainly feel useless at times (negative)	.531 (.017)	.580 (.026)	.020 (.018) ns	.005 (.027) ns
7. I feel that I'm a person of worth, at least on an equal plane with others (positive)	.596 (.019)	.359 (.042)	.073 (.018)	−.075 (.030)
8. I wish I could have more respect for myself (negative)	.392 (.020)	.390 (.031)	.123 (.02)	.033 (.031) ns
9. All in all, I am inclined to feel that I am a failure (negative)	.616 (.021)	.358 (.034)	−.017 (.023) ns	.065 (.035) ns
10. I take a positive attitude toward myself (positive)	.777 (.021)	.016 (.054) ns	.066 (.023)	−.180 (.035)

Note. ns indicates that the parameter estimate is not statistically significant at $\alpha = .05$.

the 10 factor loadings were significantly moderated. The strongest effect of individualism was found for the factor loading of Item 8: "I wish I could have more respect for myself." The factor loadings of the negative factor were not significantly moderated by individualism, while four of the five factor loadings for the positive factor were significantly larger in more individualistic countries. This corresponds with the conclusions in the original article. Individualism of the country explained 0.20% to 40.00% of the variance in the correlation coefficients across studies.

In the subgroup analysis with TSSEM, we also found a significant overall effect on the factor loadings across the groups with under versus above average scores on individualism, $\chi^2(20) = 165.773$, $p < .05$. Plotting the estimated factor loadings from the subgroup analyses without equality constraints on the loadings shows that most factor loadings for the general factor and positive factor are higher in the more individualistic samples. These plots, all other syntax, and results can be found in the online supplementary materials. The largest difference between the subgroups is found for Item 8, which matches the results of the analyses with the continuous moderator in one-stage MASEM. The factor loadings for the negative factor were very similar across subgroups.

Conclusion. By applying one-stage MASEM on the self-esteem data, we found results for the overall analyses that were nearly identical to the results obtained with TSSEM. With one-stage MASEM we were able to evaluate the regression of the factor loadings on individualism, which gives a similar, but much more precise, insight in how individualism of the samples affects how indicative the self-esteem scale is for the underlying factor.

A Simulation Study

In order to evaluate the performance of one-stage MASEM under various data conditions, we performed a simulation study. Specifically, we compared one-stage MASEM with the univariate-r approach, GLS, and TSSEM in simulated data sets with missing variables under MCAR and MNAR. We followed the design of Furlow and Beretvas (2005) who evaluated the performance of fixed-effect univariate-r approach and GLS under different patterns of missing data.

Method

Based on Furlow and Beretvas (2005), a two-factor confirmatory factor analytic (CFA) model with five indicators ($\times 1 - \times 5$) was used in the simulation study. The transposed matrices of factor loadings and factor correlations were $\begin{bmatrix} 0.7 & 0.6 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.7 & 0.6 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0.6 & 1 \end{bmatrix}$, respectively. The residual (co)variance matrix of the indicators was a diagonal matrix of $[0.51 \quad 0.64 \quad 0.20 \quad 0.51 \quad 0.64]$. These parameter values result in a model implied population correlation matrix of

$$P = \begin{bmatrix} 1 & & & & \\ 0.420 & 1 & & & \\ 0.560 & 0.480 & 1 & & \\ 0.294 & 0.252 & 0.560 & 1 & \\ 0.252 & 0.216 & 0.480 & 0.420 & 1 \end{bmatrix}.$$

Furlow and Beretvas (2005) generated multivariate normal data from P using a fixed-effects (also known as the common-effects) model. Because the present study focuses on the random-effects model, we generated heterogeneous population correlation matrices. Bosco, Aguinis, Singh, Field, and Pierce (2015) conducted a large-scale review to benchmark the typical correlations observed in applied psychology. Among other findings, they found that the range of population standard deviations (τ) of correlations between broad relation types, for example, the correlation between attitudes and intentions, ranged from .040 to .207 (Bosco et al., 2015; Table 3). Therefore, we used $\tau = .10$ as the population heterogeneity for all correlation coefficients, and P as the average population correlation matrix in this simulation study.

Furlow and Beretvas (2005) used a sample size of 100 for all studies. Because of the presence of missing data, the aggregated correlations were based on different sample sizes. They investigated four factors in their simulation study: number of studies (10, 30), percentage of studies with missingness (40% and 60%), percentage of variables missing (40% and 60%), and types of missingness (MCAR and MNAR). We adopted these settings in the present study and added a condition with 50 studies. The methods to induce missing data, which were the same as what Furlow and Beretvas (2005) used, are explained below.

Percentage of studies with missingness. Either 40% or 60% of the studies had missing data. In the condition of 10 studies, it means that four (40%) or six (60%) studies included missing data. In conditions with 30 studies, 12 (40%) or 18 (60%) of the studies had missing data. In the condition of 50 studies, 20 (40%) or 30 (60%) studies included missing data.

Percentage of variables missing. For the studies with missing data, either 40% or 60% of the variables were missing. Because there were five variables in the model, it means that either two or three variables would be missing. Suppose that x1 and x2 were selected missing, then all correlations of the other variables with x1 and x2 are missing. When there is no missing variable in a study, there are 10 observed correlation coefficients. This means that in the presence of missing variables, there were only three or one correlations observed in the conditions of 40% and 60% missing.

Types of missingness. In the condition of MCAR, all the studies and variables had an equal probability of being selected as missing. For the conditions with MNAR data, we followed the procedure of Furlow and Beretvas (2005). First, we calculated the average correlation of the variables with other variables. In the condition with two missing variables (40% missing variables), we selected x2 and x5 that could be potentially missing because they had the lowest average correlations. For the generated correlation matrices, the sums of the correlations associated with x2 and x5 were calculated and rank ordered. Studies with the smallest values on the sum were labeled as missing in both x2 and x5. This process attempts to mirror the scenario that researchers do not report on the variables with low correlations with other variables.

In the condition of three missing variables (60% missing variables), Variables x2, x4, and x5 were used to determine which studies would include missing correlations in the generated data. The sums of the correlations associated with x2, x4, and x5 were calculated and ranked. For the studies with the lowest sums, we removed variables x2, x4, and x5.

To summarize, there were 3 (number of studies) × 2 (percentage of studies with missingness) × 2 (percentage of variables missing) × 2 (types of missingness) = 24 conditions. The number of replications was 1,000. Each of these conditions was analyzed with four methods, being the univariate-r approach (with the harmonic mean across correlation coefficients as the sample size), GLS, TSSEM, and one-stage MASEM.

The usefulness of a computer simulation depends on how representative the evaluated conditions are for applied research. Rosopa and Kim (2017) summarized 94 published studies in human resource management, industrial-organizational psychology, and allied fields using MASEM. Based on their Table 1, we found that the 25th, 50th, and 75th percentiles of the numbers of variables and numbers of studies were four, six, and nine, and 25, 42, and 69, respectively. In our simulation, there were five variables and 10, 30, or 50 studies. We also calculated the aggregated sample sizes per correlation. In the study by Rosopa and Kim (2017), the 25th, 50th, and 75th percentiles were 547, 1,838, and 3,590 while the same statistics in our simulation were 700, 1,800, and 2,700. Therefore, our conditions seem to be quite representative of empirical studies using MASEM.

Procedure and analyses. Multivariate normal data of population correlation matrices were first generated from P with $\tau = .10$. If the generated correlation matrices were nonpositive definite,

they were regenerated until they were positive definite. The population correlation matrices were then used to generate sample correlation matrices with possible missing variables based on the conditions described above. These correlation matrices were analyzed by four different methods.

The test statistic, the associated p value, parameter estimates, and their standard errors were recorded for each replication. The degrees of freedom (dfs) of the proposed two-factor CFA is 3. Under the null hypothesis, and if all assumptions hold, the test statistic is expected to approximately follow a chi-square distribution with 3 dfs. We evaluated the mean of the chi-square statistics and the rejection rates at $\alpha = .05$ in all conditions. In addition, we calculated the relative percentage bias of the parameter estimate $B(\hat{\theta})$ of the six factor loadings and one factor correlation by

$$B(\hat{\theta}) = \frac{\bar{\hat{\theta}} - \theta}{\theta} \times 100\%, \qquad (8)$$

where $\bar{\hat{\theta}}$ and θ are the mean of the parameter estimates over 1,000 replications and the population parameter, respectively. We also calculated the relative percentage bias of the standard error $B(\overline{SE}(\hat{\theta}))$ of the factor loadings and factor correlation using

$$B(\overline{SE}(\hat{\theta})) = \frac{\overline{SE}(\hat{\theta}) - SD(\hat{\theta})}{SD(\hat{\theta})} \times 100\%, \qquad (9)$$

where $\overline{SE}(\hat{\theta})$ and $SD(\hat{\theta})$ are the mean of the standard errors and the empirical standard deviation of the parameter estimates over 1,000 replications, respectively. Hoogland and Boomsma (1998) recommended using a cutoff of a maximum of 5% for acceptable levels of bias in parameter estimates and maximum 10% for acceptable levels of standard error bias. All the simulations were conducted in R (R Core Team, 2019) with the metaSEM (Cheung, 2015b) and metafor (Viechtbauer, 2010) packages. The R code is available in Github (https://github.com/mikewlcheung/code-in-articles).

Results

All replications were convergent without problems (99% or more of the replications in all conditions for all methods). Tables 5 and 6 display the means of the test statistics and their rejection percentages. The test statistic of the univariate-r approach was largely overestimated in all conditions. The means of the test statistics obtained with the univariate-r approach varied from 11.62 to 19.55, which is much higher than the expected value of 3. The rejection percentages varied from 56.70% to 81.40% with $\alpha = .05$. For the other three methods (GLS, TSSEM and one-stage MASEM), when the data are MCAR, the test statistic approximately followed a chi-square distribution with 3 dfs as expected, although the rejection percentages were slightly above the nominal value of 5%. In conditions with 10 studies, GLS and TSSEM lead to inflated rejection percentages (around 11%), while one-stage MASEM performs slightly better with rejection percentages around 6%. Still, this indicates that 10 studies may generally not be sufficient to conduct a random-effects MASEM. In all conditions with MNAR, the test statistic was slightly inflated for GLS, TSSEM, and one-stage SEM, with mean test statistics around 4. The rejection percentages were highest for GLS and TSSEM, and somewhat lower for one-stage MASEM. This indicates that the presence of missing data with MNAR affects the performance of the test statistic.

Table 5
Means of the Test Statistics (a Chi-Square Test With 3 Degrees of Freedom for the Proposed Model)

Type of missing data	Number of studies	Proportion of studies with missing data	Proportion of variables with missing data	Uni_r	GLS	TSSEM	OSMASEM
MCAR	10	.4	.4	12.03	3.49	3.94	3.70
MCAR	10	.4	.6	11.78	3.59	3.90	3.80
MCAR	10	.6	.4	12.21	3.60	3.85	3.97
MCAR	10	.6	.6	11.84	3.83	3.83	4.20
MCAR	30	.4	.4	11.62	2.89	3.18	3.08
MCAR	30	.4	.6	12.19	3.14	3.40	3.31
MCAR	30	.6	.4	12.77	3.17	3.43	3.34
MCAR	30	.6	.6	12.33	3.19	3.33	3.35
MCAR	50	.4	.4	11.70	2.85	3.10	3.06
MCAR	50	.4	.6	12.04	2.99	3.22	3.18
MCAR	50	.6	.4	13.09	3.14	3.35	3.34
MCAR	50	.6	.6	12.49	3.07	3.20	3.23
MNAR	10	.4	.4	12.86	4.26	4.15	3.91
MNAR	10	.4	.6	12.57	4.40	4.16	3.79
MNAR	10	.6	.4	14.30	5.16	4.52	4.65
MNAR	10	.6	.6	12.77	4.66	3.93	4.19
MNAR	30	.4	.4	14.81	4.41	4.15	3.96
MNAR	30	.4	.6	12.67	3.83	3.51	3.42
MNAR	30	.6	.4	17.02	5.50	4.87	4.67
MNAR	30	.6	.6	13.29	4.33	3.52	3.59
MNAR	50	.4	.4	16.31	4.64	4.39	4.25
MNAR	50	.4	.6	12.47	3.65	3.31	3.28
MNAR	50	.6	.4	19.55	6.00	5.42	5.16
MNAR	50	.6	.6	13.25	4.15	3.36	3.48

Note. MCAR = missing completely at random; MNAR = missing not at random; Unir_r = univariate r; GLS = generalized least squares; TSSEM = two-stage structural equation modeling; OSMASEM = one-stage meta-analytic structural equation modeling.

Figure 8 shows the relative percentages bias of the parameter estimates under MCAR and MNAR. When the missing data were MCAR, the parameter estimates were unbiased; The relative bias was less than 3% for all methods in all MCAR conditions. However, when the missing data were MNAR, the relative percentage bias could be as large as −10% to −13%. The amount of bias in the parameter estimates was consistent across methods.

Figure 9 displays the relative percentage bias of the standard errors under MCAR and MNAR. For the univariate-r approach, the estimated standard errors were seriously underestimated (−21% to −52%) in all conditions. When the missing data were MCAR, the relative bias in the estimated standard errors of GLS, TSSEM, and one-stage MASEM was generally within ±10%, except in the condition of 10 studies. In the conditions with 10 studies, the results for GLS are still within the acceptable range for most parameters. The standard errors of one-stage MASEM and TSSEM are on the boundary of 10% but are increasingly underestimated with more missing data. When the missing data were MNAR, the relative bias in the estimated standard errors of GLS, TSSEM, and one-stage MASEM was still roughly within ±10%, with increasingly better performance with larger numbers of studies.

Conclusion and Discussion of the Simulation Study

The present simulation study provides some initial support of the proposed one-stage MASEM. Several conclusions can be drawn from the simulation study. When the missingness mechanism is MCAR, the parameter estimates of all methods are unbiased. Both GLS, TSSEM, and one-stage MASEM work well in terms of the chi-square statistic and estimated standard errors. The

univariate-r approach leads to inflated chi-square statistics and severely underestimated standard errors in all conditions, and should not be used in practice.

When the missingness mechanism is MNAR, some parameter estimates are systematically biased for all methods. For example, the estimated factor loadings of x2 and x5 are consistently overestimated because the studies with the smallest average correlations related to x2 and x5 are systematically missing in the simulation study. Although some parameter estimates are biased, the estimated standard errors seem to work reasonably well, except for the univariate-r approach.

To summarize, both GLS, TSSEM, and one-stage MASEM work reasonably well under missing data of MCAR and at least 30 studies. The performance of all methods is affected by the presence of missing data of MNAR. These findings are consistent with earlier studies of missing data in meta-analysis (Furlow & Beretvas, 2005; Mavridis, Chaimani, Efthimiou, Leucht, & Salanti, 2014), and underscore the importance of avoiding data MNAR by including as much as possible unpublished results in a meta-analysis.

We evaluated existing and the new MASEM approach in several realistic scenarios based on the reviews of meta-analyses and MASEM studies by Bosco et al. (2015) and Rosopa and Kim (2017). However, we did not evaluate all possible conditions. First, we have only analyzed one CFA model, which was based on the study by Furlow and Beretvas (2005). Obviously, there are other CFA models, full SEM models, and path models with different model complexity that would also be interesting to consider. Second, we did not evaluate models with moderated effects in the

Table 6
Rejection Percentages (α = .05) for Testing the Proposed Model With df = 3

Type of missing data	Number of studies	Proportion of studies with missing data	Proportion of variables with missing data	Uni_r	GLS	TSSEM	OSMASEM
MCAR	10	.4	.4	60.00	9.90	12.70	6.50
MCAR	10	.4	.6	57.30	10.50	11.21	5.30
MCAR	10	.6	.4	56.70	10.60	11.50	6.30
MCAR	10	.6	.6	58.26	12.30	11.03	6.80
MCAR	30	.4	.4	57.36	4.10	6.50	5.00
MCAR	30	.4	.6	59.40	6.80	8.20	7.10
MCAR	30	.6	.4	59.60	7.20	8.30	6.90
MCAR	30	.6	.6	59.30	6.80	7.40	6.70
MCAR	50	.4	.4	57.70	5.30	6.20	6.00
MCAR	50	.4	.6	56.70	5.00	6.30	5.60
MCAR	50	.6	.4	62.30	5.80	7.40	6.20
MCAR	50	.6	.6	59.10	5.60	6.50	6.20
MNAR	10	.4	.4	63.00	14.51	12.00	5.50
MNAR	10	.4	.6	59.90	15.60	14.80	6.80
MNAR	10	.6	.4	66.60	20.92	15.00	7.21
MNAR	10	.6	.6	57.56	18.05	12.30	6.40
MNAR	30	.4	.4	68.10	16.00	13.40	11.80
MNAR	30	.4	.6	61.50	10.10	7.50	6.40
MNAR	30	.6	.4	71.80	24.60	19.90	17.20
MNAR	30	.6	.6	62.30	14.90	9.40	7.40
MNAR	50	.4	.4	71.90	17.10	15.80	13.90
MNAR	50	.4	.6	61.60	9.00	6.30	5.40
MNAR	50	.6	.4	81.40	25.90	20.70	18.00
MNAR	50	.6	.6	60.50	13.90	7.50	8.00

Note. MCAR = missing completely at random; MNAR = missing not at random; Unir_r = univariate r; GLS = generalized least squares; TSSEM = two-stage structural equation modeling; OSMASEM = one-stage meta-analytic structural equation modeling.

simulation study. This choice enabled us to compare the different MASEM method to each other. A more comprehensive simulation study evaluating one-Stage MASEM with continuous moderators in a wide array of situations would be desirable for future research. Third, we have manipulated data with MNAR following the operationalization defined by Furlow and Beretvas (2005). Although we think this is a reasonable procedure, there are other ways to operationalize missing data with MNAR. For example, researchers may decide not to report some variables if the associated regression coefficients instead of correlation coefficients are nonsignificant or small. Also, in the case of MNAR data, it may actually be more realistic if the percentage of missing data is smaller than 40% or 60%. It is reasonable to expect that missing variables in MASEM are a mix of MCAR and MNAR. Some variables are not measured in the primary studies because these researchers are not interested in these constructs. In some studies, researchers have measured these variables and chosen not to report them because they are nonsignificant. Future research may explore how different methods perform under alternative MNAR conditions.

General Discussion

Summary

We presented one-stage MASEM as a novel method that allows for the moderation of SEM-parameters by all types of moderators, including continuous moderators. With one-stage MASEM, researchers can evaluate their research questions about average parameters as well as differences in parameters across studies within the same model. Another strength of this approach is that there are

readily available R-functions to apply the method. The specification of the structural equation model, which has been a source of trouble when researchers had to specify the model in matrix form, has been improved since the metaSEM package may formulate models with the lavaan syntax (Rosseel, 2012). While previously, researchers had to define several large matrices specifying each fixed and free parameter, one can now formulate any model using only some simple operators, parameter labels, and variable names.

This article served to introduce the new one-stage MASEM, evaluate its performance in a simulation study, and illustrate its usefulness with real data. We will provide some practical recommendations for researchers who would like to apply one-stage MASEM in their research in the next section. Next, we will discuss some potentially interesting avenues for future research.

Practical Recommendations for the Applications of the One-Stage MASEM

When applying random-effects MASEM in general, and one-stage MASEM in specific, one may encounter some practical issues. Our simulation study showed that the number of studies in the MASEM should be at least 30 in order to obtain valid results. However, even if the number of included studies is larger than 30, the computer program may sometimes have difficulties to converge to an optimal solution. The functions implemented in metaSEM will, in this case, give a warning that the OpenMx status is not either 0 or 1. One option to obtain convergence is to increase the number of attempts of fitting the model using the rerun() function. Applying this function will rerun the model until the optimizer yields an acceptable solution or until the maximum
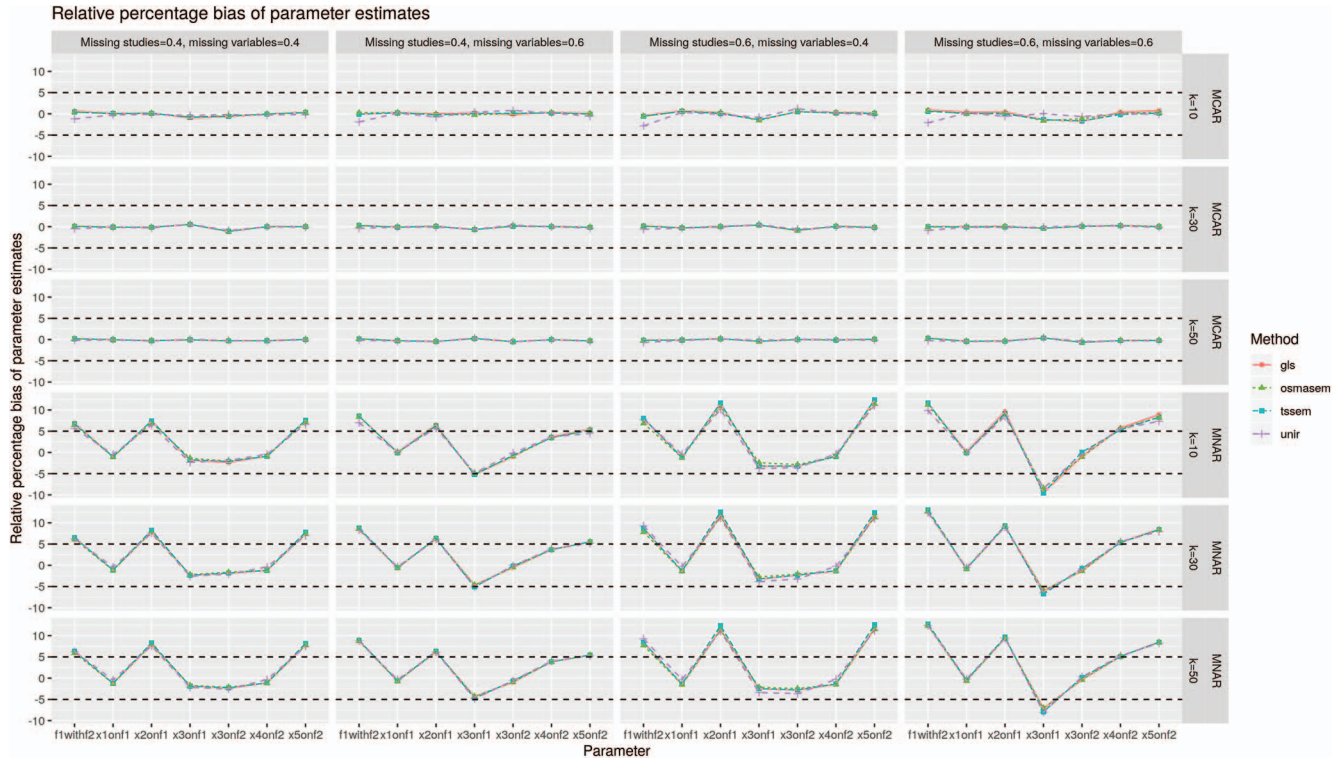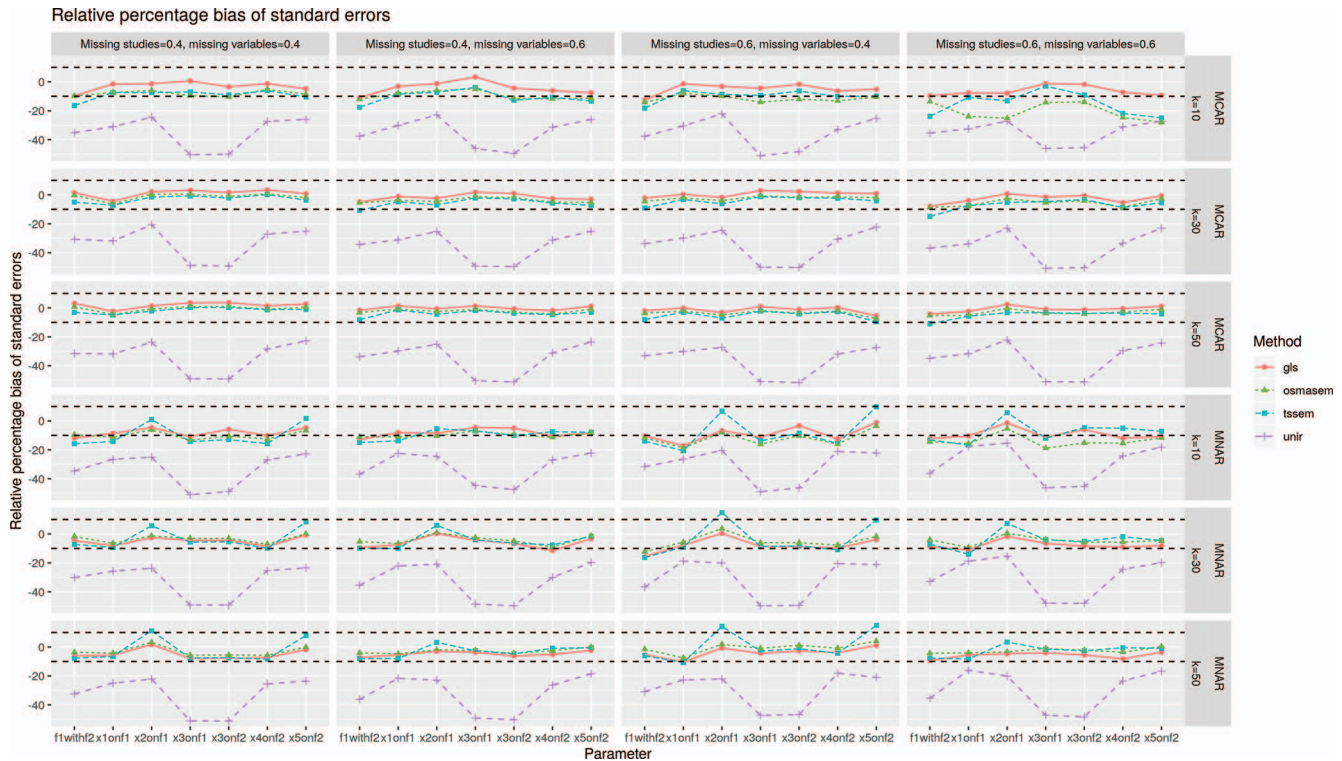
Figure 8. Relative percentage bias of parameter estimates. k = number of studies; MCAR = missing completely at random; MNAR = missing not at random; missing studies = 0.4; missing variables = 0.6; percentage of studies with missingness = 40, and percentage of missing variables = 60%; Unir_r = univariate r; GLS = generalized least squares; TSSEM = two-stage structural equation modeling; OSMASEM = one-stage meta-analytic structural equation modeling; f1 with f2 = correlation between the two latent factors; x1 on f1 = factor loading of x1 on f1. See the online article for the color version of this figure.

number of attempts is reached. Between attempts, new start values are randomly perturbed, and optimization-control parameters may be altered (Neale et al., 2016). Another simple strategy is to center or standardize the continuous moderators. It will improve the numerical stability of the results and increase the chance of convergence.

Estimation and convergence difficulties quite frequently relate to the estimation of the between-studies (co)variances in $\mathbf{T}^2$. The dimensions of this matrix can become very large with increasing numbers of variables. For example, with five observed variables in the MASEM model, there will be $5 \times 4/2 = 10$ correlations between these variables. The covariance matrix of these 10 correlations will thus contain $10 \times 11/2 = 55$ unique elements. Because the number of studies in MASEM is often smaller than 55, it is almost never possible to obtain reliable estimates of both the between-studies variances and covariances. Therefore, it is recommended to fix the between-studies covariances to zero, meaning that $\mathbf{T}^2$ will be diagonal (Becker, 2009). In the one-stage MASEM functions in metaSEM, the default model specification restricts $\mathbf{T}^2$ to be diagonal. Note that this restriction does not mean that the observed correlations are assumed to be independent because the within-studies covariance matrices ($\mathbf{V}_i$) contain off-diagonal elements representing sampling covariance. A diagonal $\mathbf{T}^2$ implies that the population correlation coefficients are independent at the between-study level.

Even with a diagonal $\mathbf{T}^2$, one may encounter problems with estimation of the between-studies variances. For example, the between-studies variance for one of the correlation coefficients may be very small or zero, leading to a negative estimate, or an estimate at the lower bound of zero. To avoid negative estimates of variances, the one-stage MASEM functions in metaSEM apply either an exponential-log transformation on the variances or squared standard deviations. Still, the real between-studies variance of coefficients may be zero, or the number of studies contributing an effect size for a specific correlation coefficient may be relatively small, leading to overparameterized models and associated estimation issues. It may then be necessary to fix some individual variances to zero. This can be done using an optional argument of the osmasem()-function.

In some cases, adding moderators to one-stage MASEM may result in a negative R^2 associated with the τ^2's, implying that the between-studies variance increased when adding moderators to the model. This was actually the case in our illustration with the Nohe-data for two of the six correlations. If this happens for variable pairs that are not connected by a moderated parameter, and the R^2 is small, the increase in between-studies variance may be practically zero, and the R^2 is truncated to 0 (e.g., Borenstein et al., 2009). This may, for example, be encountered if the number of studies is small, or if there are many missing effect sizes for a specific relationship that is tested for moderation. In these cases,

Relative percentage bias of standard errors

Figure 9. Relative percentage bias of standard errors of parameter estimates. k = number of studies; MCAR = missing completely at random; MNAR = missing not at random; missing studies = 0.4; missing variables = 0.6; percentage of studies with missingness = 40, and percentage of missing variables = 60%; Unir_r = univariate r; GLS = generalized least squares; TSSEM = two-stage structural equation modeling; OSMASEM = one-stage meta-analytic structural equation modeling; f1 with f2 = correlation between the two latent factors; x1 on f1 = factor loading of x1 on f1. See the online article for the color version of this figure.

the estimation of the between-studies variance is unstable, leading to an unstable estimate of R^2. Unstable estimates of the between-studies variance lower number of studies are not specific to MASEM, as this phenomenon is well-known in the standard meta-analysis (Chung, Rabe-Hesketh, & Choi, 2013; Sidik & Jonkman, 2007; Veroniki et al., 2016).

Another issue is the occurrence of inadmissible parameter estimates, such as correlation coefficients that are larger than one, or many negative R^2-values. Just as in standard structural equation modeling such a finding is likely indicative of either a misspecified or overparameterized model (Chen, Bollen, Paxton, Curran, & Kirby, 2001). In such cases, researchers could evaluate how the model can be respecified sensibly in order to decrease the number of parameters to be estimated.

We have presented a method to evaluate moderation effects on model parameters in MASEM, in which there is no theoretical limit to the number of moderator variables that can be analyzed. In practice, the number of moderator variables that can be added to the model is limited, because the number of studies included in MASEM is typically smaller than in standard meta-analysis. Reviews of applied MASEM studies indicated that a number of around 30 studies is typical in MASEM (Sheng et al., 2016). In our examples, we found that convergence was difficult to obtain when analyzing more than one moderator. It is therefore recommended that researchers carefully select the moderator variables of interest.

With standard meta-analysis, researchers sometimes include a large number of moderator variables for which there are no a priori hypotheses formulated (Li, Dusseldorp, & Meulman, 2017; Song, Sheldon, Sutton, Abrams, & Jones, 2001). One-stage MASEM is not suited for such exploratory analyses. We recommend to only evaluate moderator variables that are related to the main research question(s). In our illustration with the self-esteem data, for example, the authors expected the scale to be more indicative in individualistic countries because expressions of overly positive self-views are typically seen as less appropriate in less individualistic societies (Gnambs, Scharl, & Schroeders, 2018). Other study-level moderators such as the percentage of males in the sample and the mean age of the sample were also available, but were not related to the research question and therefore not analyzed.

Even if the number of studies in a MASEM-analysis is large, the number of observed effect sizes for a particular relation in the model may be small. For example, many studies may report correlations between variables like attitudes and behavioral intention, while there are only a limited number of correlations reported for the relations between attitudes and actual behavior because the actual behavior is much harder to measure than behavioral intentions. As a result, it may be possible to look at moderator effects on the regression of behavioral intention on attitudes, but not to evaluate moderator effects on the regression coefficients concern-

Comparisons of approaches via simulations. Besides the frequentist approaches listed in Table 1, Ke et al. (2018) recently proposed a Bayesian approach to conduct MASEM. Their approach can be extended to include categorical and continuous moderators. It is of interest to compare the empirical performance of these approaches via computer simulations. However, it is not straightforward to design a simulation that fits all models in correlation- and parameter-based MASEM. Several issues have to be carefully considered. One of them is the source of heterogeneity. The correlation-based MASEM, including the univariate-r, GLS, TSSEM, and one-stage MASEM approaches assume that the variation (heterogeneity) is due to the correlation coefficients. On the other hand, the FIMASEM, parameter-based MASEM, and Bayesian MASEM theorize that the variation in the SEM parameters is the source of heterogeneity. The correlation coefficients and the parameters in SEM are nonlinearly related. Therefore, if researchers generate data from one model, say the correlation-based MASEM, and fit the data with the other model, say the parameter-based MASEM, the model is misspecified (Cheung, Sun, & Chan, 2019). One possible solution is to generate data from both models and see how robust these approaches are. Other interesting factors to consider for the simulation studies could, for example, be the number of studies, the sample sizes of the individual studies, the patterns of missing data, models with and without a moderator, the number of moderators, the number of parameters to be moderated, and the employed moderating functions.

Conclusion

We proposed a new MASEM method that is better suitable to explain study-level heterogeneity than existing methods. We illustrated the method using two real data examples and provided initial insight in its empirical performance via a computer simulation. We discussed several practical issues about the method, and provided user-friendly functions in the metaSEM package, facilitating the use of one-stage MASEM in substantive research.

References

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management, 37,* 5–38. http://dx.doi.org/10.1177/0149206310377113

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Aloe, A. M. (2014). An empirical investigation of partial effect sizes in meta-analysis of correlational data. *The Journal of General Psychology, 141,* 47–64. http://dx.doi.org/10.1080/00221309.2013.853021

Aloe, A. M., & Becker, B. J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics, 37,* 278–297. http://dx.doi.org/10.3102/1076998610396901

Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods, 21,* 1–12. http://dx.doi.org/10.1037/met0000051

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22,* 507–526. http://dx.doi.org/10.1037/met0000077

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Tradi-

tional and new models. *Psychological Methods, 14,* 101–125. http://dx.doi.org/10.1037/a0015583

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics, 17,* 341–362. http://dx.doi.org/10.3102/10769986017004341

Becker, B. J. (1995). Corrections to "Using results from replicated studies to estimate linear models". *Journal of Educational and Behavioral Statistics, 20,* 100–102.

Becker, B. J. (2007). Multivariate meta-analysis: Contributions of Ingram Olkin. *Statistical Science, 22,* 401–406. http://dx.doi.org/10.1214/07-STS239

Becker, B. J. (2009). Model-based meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 377–395). New York, NY: Russell Sage Foundation.

Becker, B. J., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science, 22,* 414–429. http://dx.doi.org/10.1214/07-STS243

Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P., . . . Joo, H. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal, 37,* 477–497. http://dx.doi.org/10.1002/smj.2338

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine, 17,* 2537–2550. http://dx.doi.org/10.1002/(SICI)1097-0258(19981130)17:22<2537::AID-SIM953>3.0.CO;2-C

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9780470743386

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100,* 431–449. http://dx.doi.org/10.1037/a0038047

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29,* 468–508. http://dx.doi.org/10.1177/0049124101029004003

Cheung, M. W.-L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods, 13,* 182–202. http://dx.doi.org/10.1037/a0013163

Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling, 20,* 429–454. http://dx.doi.org/10.1080/10705511.2013.797827

Cheung, M. W.-L. (2014a). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods, 46,* 29–40. http://dx.doi.org/10.3758/s13428-013-0361-y

Cheung, M. W.-L. (2014b). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19,* 211–229. http://dx.doi.org/10.1037/a0032968

Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach.* West Sussex, UK: Wiley.

Cheung, M. W.-L. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology.* Advance online publication. http://dx.doi.org/10.3389/fpsyg.2014.01521

Cheung, M. W.-L. (2018). Issues in solving the problem of effect size heterogeneity in meta-analytic structural equation modeling: A commentary and simulation study on Yu, Downes, Carter, and O'Boyle (2016). *Journal of Applied Psychology, 103,* 787–803. http://dx.doi.org/10.1037/apl0000284

Cheung, M. W.-L. (2019). Some reflections on combining meta-analysis and structural equation modeling. *Research Synthesis Methods, 10,* 15–22. http://dx.doi.org/10.1002/jrsm.1321

Cheung, M. W.-L., & Chan, W. (2004). Testing dependent correlation coefficients via structural equation modeling. *Organizational Research Methods, 7,* 206–223. http://dx.doi.org/10.1177/1094428104264024

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10,* 40–64. http://dx.doi.org/10.1037/1082-989X.10.1.40

Cheung, M. W.-L., & Cheung, S. F. (2016). Random-effects models for meta-analytic structural equation modeling: Review, issues, and illustrations. *Research Synthesis Methods, 7,* 140–155. http://dx.doi.org/10.1002/jrsm.1166

Cheung, M. W.-L., & Jak, S. (2017, July). *A simulation study on meta-analytic structural equation modeling under random-effects model.* Paper presented at the 12th Annual Meeting of the Society for Research Synthesis Methodology, Montreal, Québec, Canada.

Cheung, S. F., Sun, R. W., & Chan, D. K.-S. (2019). Correlation-based meta-analytic structural equation modeling: Effects of parameter covariance on point and interval estimates. *Organizational Research Methods, 22,* 892–916. http://dx.doi.org/10.1177/1094428118770736

Chung, Y., Rabe-Hesketh, S., & Choi, I.-H. (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine, 32,* 4071–4089. http://dx.doi.org/10.1002/sim.5821

Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles, CA: Sage.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327. http://dx.doi.org/10.1037/0033-2909.105.2.317

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14,* 81–100. http://dx.doi.org/10.1037/a0015914

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89–98.

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford Press.

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17,* 120–128. http://dx.doi.org/10.1037/a0024445

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron, 1,* 3–32.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate non-normality. *Structural Equation Modeling, 7,* 356–410. http://dx.doi.org/10.1207/S15328007SEM0703_2

Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods, 10,* 227–254. http://dx.doi.org/10.1037/1082-989X.10.2.227

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733–760.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8. http://dx.doi.org/10.3102/0013189X005010003

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York, NY: Russell Sage Foundation.

Gnambs, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg Self-Esteem Scale. *Zeitschrift fur Psychologie, 226,* 14–29. http://dx.doi.org/10.1027/2151-2604/a000317

Gnambs, T., & Staufenbiel, T. (2016). Parameter accuracy in meta-analyses of factor structures. *Research Synthesis Methods, 7,* 168–186. http://dx.doi.org/10.1002/jrsm.1190

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

Hedges, L. V. (2016). Applying meta-analysis to structural equation modeling. *Research Synthesis Methods, 7,* 209–214. http://dx.doi.org/10.1002/jrsm.1214

Hedges, L. V., & Schauer, J. M. (2018). Statistical analyses for studying replication: Meta-analytic perspectives. *Psychological Methods.* Advance online publication. http://dx.doi.org/10.1037/met0000189

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39–65. http://dx.doi.org/10.1002/jrsm.5

Hedges, L. V., & Vevea, J. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. http://dx.doi.org/10.1037/1082-989X.3.4.486

Hong, R. Y., & Cheung, M. W. L. (2015). The structure of cognitive vulnerabilities to depression and anxiety: Evidence for a common core etiologic process based on a meta-analytic review. *Clinical Psychological Science, 3,* 892–912. http://dx.doi.org/10.1177/2167702614553789

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling an overview and a meta-analysis. *Sociological Methods & Research, 26,* 329–367. http://dx.doi.org/10.1177/0049124198026003003

Hu, L.-T., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Jackson, D., Riley, R., & White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine, 30,* 2481–2498. http://dx.doi.org/10.1002/sim.4247

Jackson, D., White, I. R., & Thompson, S. G. (2010). Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine, 29,* 1282–1297. http://dx.doi.org/10.1002/sim.3602

Jak, S. (2015). *Meta-analytic structural equation modelling.* Amsterdam, the Netherlands: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-27174-3

Jak, S., & Cheung, M. W.-L. (2018a). Accounting for missing correlation coefficients in fixed-effects MASEM. *Multivariate Behavioral Research, 53,* 1–14. http://dx.doi.org/10.1080/00273171.2017.1375886

Jak, S., & Cheung, M. W.-L. (2018b). Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis. *Behavior Research Methods, 50,* 1359–1373. http://dx.doi.org/10.3758/s13428-018-1046-3

Jamshidian, M., & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24,* 21–24. http://dx.doi.org/10.3102/10769986024001021

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23,* 524–532.

Jones, A. P., Riley, R. D., Williamson, P. R., & Whitehead, A. (2009). Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clinical Trials, 6,* 16–27. http://dx.doi.org/10.1177/1740774508100984

Ke, Z., Zhang, Q., & Tong, X. (2018). Bayesian meta-analytic SEM: A one-stage approach to modeling between-studies heterogeneity in structural parameters. *Structural Equation Modeling.* Advance online publication. http://dx.doi.org/10.1080/10705511.2018.1530059

Kisamore, J. L., & Brannick, M. T. (2008). An illustration of the consequences of meta-analysis model choice. *Organizational Research Methods, 11,* 35–53. http://dx.doi.org/10.1177/1094428106287393

Klein, A., & Muthén, B. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42,* 647–673. http://dx.doi.org/10.1080/00273170701710205

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2,* 61–76. http://dx.doi.org/10.1002/jrsm.35

Lee, S.-Y., & Zhu, H.-T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika, 67,* 189–210. http://dx.doi.org/10.1007/BF02294842

Li, X., Dusseldorp, E., & Meulman, J. J. (2017). Meta-CART: A tool to identify interactions between moderators in meta-analysis. *British Journal of Mathematical & Statistical Psychology, 70,* 118–136. http://dx.doi.org/10.1111/bmsp.12088

Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling, 13,* 497–519. http://dx.doi.org/10.1207/s15328007sem1304_1

Loehlin, J. C., & Beaujean, A. A. (2016). *Latent variable models: An introduction to factor, path, and structural equation analysis.* London, UK: Taylor & Francis.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149. http://dx.doi.org/10.1037/1082-989X.1.2.130

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40. http://dx.doi.org/10.1037/1082-989X.7.1.19

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9,* 275–300. http://dx.doi.org/10.1037/1082-989X.9.3.275

Mavridis, D., Chaimani, A., Efthimiou, O., Leucht, S., & Salanti, G. (2014). Addressing missing outcome data in meta-analysis. *Evidence-Based Mental Health, 17,* 85–89. http://dx.doi.org/10.1136/eb-2014-101900

McArdle, J. J. (2005). The development of the RAM rules for latent variable structural equation modeling. In A. Maydeau-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A Festschrift for Roderick P. McDonald* (pp. 225–274). Mahwah, NJ: Erlbaum.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the RAM logic for structural equation model specification. *British Journal of Mathematical & Statistical Psychology, 37,* 234–251. http://dx.doi.org/10.1111/j.2044-8317.1984.tb00802.x

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10,* 259–284. http://dx.doi.org/10.1037/1082-989X.10.3.259

Minkov, M., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., . . . Mudd, B. (2017). A revision of Hofstede's individualism-collectivism dimension: A new national index from a 56-country study. *Cross Cultural & Strategic Management, 24,* 386–404. http://dx.doi.org/10.1108/CCSM-11-2016-0197

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132. http://dx.doi.org/10.1007/BF02294210

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22,* 376–398. http://dx.doi.org/10.1177/0049124194022003006

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431–462. http://dx.doi.org/10.1007/BF02294365

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., . . . Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika, 81,* 535–549. http://dx.doi.org/10.1007/s11336-014-9435-8

Nohe, C., Meier, L. L., Sonntag, K., & Michel, A. (2015). The chicken or the egg? A meta-analysis of panel studies of the relationship between work-family conflict and strain. *Journal of Applied Psychology, 100,* 522–536. http://dx.doi.org/10.1037/a0038012

Norton, S., Cosco, T., Doyle, F., Done, J., & Sacker, A. (2013). The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research, 74,* 74–81. http://dx.doi.org/10.1016/j.jpsychores.2012.10.010

Olkin, I., & Siotani, M. (1976). Asymptotic distribution of functions of a correlation matrix. In S. Ideka (Ed.), *Essays in probability and statistics* (pp. 235–251). Tokyo, Japan: Shinko Tsusho.

Oort, F. J., & Jak, S. (2016). Maximum likelihood estimation in meta-analytic structural equation modeling. *Research Synthesis Methods, 7,* 156–167. http://dx.doi.org/10.1002/jrsm.1203

Pigott, T. D. (2009). Handling missing data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 399–416). New York, NY: Russell Sage Foundation.

R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400876136

Rosopa, P. J., & Kim, B. (2017). Robustness of statistical inferences using linear models with meta-analytic correlation matrices. *Human Resource Management Review, 27,* 216–236. http://dx.doi.org/10.1016/j.hrmr.2016.09.012

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York, NY: Wiley and Sons. http://dx.doi.org/10.1002/9780470316696

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. http://dx.doi.org/10.1037/1082-989X.7.2.147

Schmid, C. H. (2001). Using Bayesian inference to perform meta-analysis. *Evaluation & the Health Professions, 24,* 165–189. http://dx.doi.org/10.1177/01632780122034867

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical & Statistical Psychology, 62,* 97–128. http://dx.doi.org/10.1348/000711007X255327

Schulze, R. (2007). Current methods for meta-analysis. *The Journal of Psychology, 215,* 90–103.

Shadish, W. R. (1996). Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods, and issues. *Psychological Methods, 1,* 47–65. http://dx.doi.org/10.1037/1082-989X.1.1.47

Sheng, Z., Kong, W., Cortina, J. M., & Hou, S. (2016). Analyzing matrices of meta-analytic correlations: Current practices and recommendations. *Research Synthesis Methods, 7,* 187–208. http://dx.doi.org/10.1002/jrsm.1206

Sidik, K., & Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine, 26,* 1964–1981. http://dx.doi.org/10.1002/sim.2688

Song, F., Sheldon, T. A., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2001). Methods for exploring heterogeneity in meta-analysis. *Evaluation & the Health Professions, 24,* 126–151. http://dx.doi.org/10.1177/016327870102400203

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B: Methodological, 64,* 583–639. http://dx.doi .org/10.1111/1467-9868.00353

Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research, 10,* 277–303. http://dx.doi.org/10.1177/096228020101000404

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20,* 375–393. http://dx .doi.org/10.1037/met0000011

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40,* 604– 634. http://dx.doi.org/10.3102/1076998615606099

Tong, X., Zhang, Z., & Yuan, K.-H. (2014). Evaluation of test statistics for robust structural equation modeling with nonnormal missing data. *Structural Equation Modeling, 21,* 553–565. http://dx.doi.org/10.1080/ 10705511.2014.919820

Topa, G., & Moriano, J. A. (2010). Theory of planned behavior and smoking: Meta-analysis and SEM model. *Substance Abuse and Rehabilitation, 1,* 23–33. http://dx.doi.org/10.2147/SAR.S15168

Ueno, T., Fastrich, G. M., & Murayama, K. (2016). Meta-analysis to integrate effect sizes within an article: Possible misuse and Type I error inflation. *Journal of Experimental Psychology: General, 145,* 643–654. http://dx.doi.org/10.1037/xge0000159

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45,* 576–594. http://dx.doi.org/10 .3758/s13428-012-0261-6

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., . . . Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods, 7,* 55–79. http://dx.doi.org/10.1002/jrsm.1164

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428–443. http://dx.doi.org/10.1037/1082-989X.10.4 .428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://dx.doi.org/10 .18637/jss.v036.i03

Visweswaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48,* 865–885. http://dx.doi.org/10.1111/j.1744-6570.1995 .tb01784.x

Wilson, S. J., Polanin, J. R., & Lipsey, M. W. (2016). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods, 7,* 121–139. http://dx.doi.org/10.1002/jrsm.1199

Yu, J. J., Downes, P. E., Carter, K. M., & O'Boyle, E. H. (2016). The problem of effect size heterogeneity in meta-analytic structural equation modeling. *Journal of Applied Psychology, 101,* 1457–1473. http://dx.doi .org/10.1037/apl0000141

Yu, J. J., Downes, P. E., Carter, K. M., & O'Boyle, E. (2018). The heterogeneity problem in meta-analytic structural equation modeling (MASEM) revisited: A reply to Cheung. *Journal of Applied Psychology, 103,* 804–811. http://dx.doi.org/10.1037/apl0000328

Yuan, K. H., & Kano, Y. (2018). Meta-analytical SEM: Equivalence between maximum likelihood and generalized least squares. *Journal of Educational and Behavioral Statistics, 43,* 693–720. http://dx.doi.org/10 .3102/1076998618787799

Zeegers, M. A. J., Colonnesi, C., Stams, G. J. M., & Meins, E. (2017). Mind matters: A meta-analysis on parental mentalization and sensitivity as predictors of infant-parent attachment. *Psychological Bulletin, 143,* 1245–1272. http://dx.doi.org/10.1037/bul0000114

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!