Maximum likelihood estimation in meta-analytic structural equation modeling

Frans J. Oort[a]* and Suzanne Jak[b,c]

Meta-analytic structural equation modeling (MASEM) involves fitting models to a common population correlation matrix that is estimated on the basis of correlation coefficients that are reported by a number of independent studies. MASEM typically consist of two stages. The method that has been found to perform best in terms of statistical properties is the two-stage structural equation modeling, in which maximum likelihood analysis is used to estimate the common correlation matrix in the first stage, and weighted least squares analysis is used to fit structural equation models to the common correlation matrix in the second stage. In the present paper, we propose an alternative method, ML MASEM, that uses ML estimation throughout. In a simulation study, we use both methods and compare chi-square distributions, bias in parameter estimates, false positive rates, and true positive rates. Both methods appear to yield unbiased parameter estimates and false and true positive rates that are close to the expected values. ML MASEM parameter estimates are found to be significantly less bias than two-stage structural equation modeling estimates, but the differences are very small. The choice between the two methods may therefore be based on other fundamental or practical arguments. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: meta-analysis; structural equation modeling; simulation study; maximum likelihood; weighted least squares

1. Introduction

The combination of meta-analysis and structural equation modeling (SEM) for the purpose of testing hypothesized models is called meta-analytic structural equation modeling or MASEM (Cheung and Chan, 2005a). Using MASEM, information from multiple studies can be used to test a single model that explains the relationships between a set of variables or to compare several models that are supported by different studies or theories. MASEM typically consists of two stages (Viswesvaran and Ones, 1995). In the first stage, correlation matrices are tested for homogeneity across studies and combined together to form a pooled correlation matrix. Several methods to synthesize correlation matrices have been proposed and evaluated (e.g., Becker, 1992, 1995, 2009; Cheung and Chan, 2005a; Hafdahl, 2007, 2008). In the second stage, a structural equation model is fitted to the pooled correlation matrix. For this stage, also several methods have been proposed (see Zhang (2011) for an overview). The method that is generally found to perform best in terms of statistical properties and flexibility is the two-stage structural equation modeling (TSSEM) approach of Cheung and Chan (2005a). In this method, maximum likelihood (ML) estimation is used to estimate a pooled correlation matrix (Stage 1), and weighted least squares (WLS) estimation is used to fit structural equation models to this correlation matrix (Stage 2). The purpose of the present paper is to propose and evaluate an alternative way of conducting MASEM, using ML estimation in all steps of the procedure.

[a]*Research Institute of Child Development and Education, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands*
[b]*Utrecht University, Utrecht, The Netherlands*
[c]*National University of Singapore, Singapore*
*Correspondence to: F. J. Oort, Research Institute of Child Development and Education, Faculty of Social and Behavioural Sciences, University of Amsterdam, P.O. Box 15780, 1001 NG Amsterdam, The Netherlands.
E-mail: f.j.oort@uva.nl*

1.1. Two-stage structural equation modeling

In Stage 1 of TSSEM, correlation matrices are tested for homogeneity across studies using multigroup modeling, after which they are combined to form a pooled correlation matrix. In Stage 2, the pooled correlation matrix is taken as the observed matrix in a SEM analysis. Cheung and Chan (2005a) use ML to estimate the pooled correlation matrix in the first stage and WLS to estimate model parameters in the second stage. In Stage 2 WLS estimation, the inversed matrix of asymptotic variances and covariances of the Stage 1 correlation estimates are used as the weight matrix in the WLS optimization function. This ensures that correlation coefficients that are estimated with more precision (for example, coefficients that are based on more or larger studies) in Stage 1 obtain more weight in the estimation of model parameters in Stage 2.

Cheung and Chan (2005a) conducted a large simulation study in which the performance of TSSEM was compared with two univariate approaches and one multivariate approach. In the univariate approaches, the correlation elements are pooled separately across studies based on bivariate information only. Cheung and Chan (2005a) use the term univariate-r method to refer to the approach from Hunter and Schmidt (2004), which involves pooling the untransformed correlation coefficients. The univariate-z method refers to the method of Hedges and Olkin (1985), who proposed using Fisher's z-transformed correlation coefficients. The multivariate approach with which they compared TSSEM is the generalized least squares (GLS) method (Becker, 1992, 1995, 2009). The simulation showed that the GLS method rejects homogeneity of correlation matrices too often and leads to biased parameter estimates at Stage 2. The univariate methods lead to inflated Type 1 errors, while TSSEM leads to unbiased parameter estimates and false positive rates close to the expected rates. The statistical power to reject an underspecified factor model was extremely high for all four methods. TSSEM thus came out as best out of these methods. A software to apply TSSEM is readily available in the R-Package metaSEM (Cheung, 2015a), which relies on the OpenMx package (Boker *et al.*, 2011) to fit the Stages 1 and 2 models. In this way, metaSEM provides parameter estimates with standard errors, a chi-square measure of fit, and likelihood-based confidence intervals for parameters at both stages of the analysis.

1.2. Meta-analytic structural equation modeling with maximum likelihood estimation

In TSSEM, the pooled correlation matrix is treated as being observed, while the precision of the estimated correlation coefficients is taken into account by using WLS estimation to fit a structural equation model in Stage 2. Instead, the structural equation model can be fitted to the observed correlations directly, by imposing restrictions on the Stage 1 model. In this way, the structural equation model is nested in the Stage 1 model, and both models can be fitted through ML estimation. This is an attractive approach because it provides an ML chi-square to evaluate the overall goodness-of-fit of the structural equation model and a chi-square difference test to evaluate the significance of the difference in fit between the two models. We will refer to this method as maximum likelihood meta-analytic structural equation modeling (ML MASEM). In the next sections, we describe TSSEM as a two-stage approach and ML MASEM as a three-step approach to MASEM. The first two steps of ML MASEM are similar to the first stage of TSSEM, as will be explained after introducing the two methods. Subsequently, we will compare the performance of TSSEM and ML MASEM in a simulation study.

2. Two-stage structural equation modeling

Stage 1: *Testing homogeneity of correlation matrices*

Let $\mathbf{R}_g$ be the $p_g \times p_g$ sample correlation matrix and p_g be the number of observed variables in the gth study. Not all studies include all variables. For example, in a meta-analysis of three variables, the correlation matrices in the first three studies may look like this:

$$\mathbf{R}_1 = \begin{bmatrix} 1 & & \\ r_{21\_1} & 1 & \\ r_{31\_1} & r_{32\_1} & 1 \end{bmatrix}, \quad \mathbf{R}_2 = \begin{bmatrix} 1 & \\ r_{32\_2} & 1 \end{bmatrix}, \quad \text{and } \mathbf{R}_3 = \begin{bmatrix} 1 & \\ r_{21\_3} & 1 \end{bmatrix}.$$

Here, Study 1 contains all variables, Study 2 misses Variable 1, and Study 3 misses Variable 3. An estimate of the population correlation matrix $\mathbf{P}$ of all q variables is obtained by fitting a multigroup SEM model, where the model for each group (study) is

$$\Sigma_g = \mathbf{D}_g \left(\mathbf{M}_g \, \mathbf{P} \, \mathbf{M}_g{}^\mathsf{T} \right) \mathbf{D}_g. \tag{1}$$

In this model, $\mathbf{P}$ is the $q \times q$ population correlation matrix with $\text{diag}(\mathbf{P}) = \mathbf{I}$, matrix $\mathbf{M}_g$ is a $p_g \times q$ selection matrix that accommodates smaller correlations matrices from studies with missing variables ($p_g < q$), and $\mathbf{D}_g$ is a $p_g \times p_g$ diagonal matrix that accounts for differences in scaling of the variables across the G studies. Setting $\mathbf{D}_g$ equal across studies ($\mathbf{D}_g = \mathbf{D}$) would imply equality of covariance matrices rather than correlation matrices (Cheung and Chan, 2009). Correct parameter estimates can be obtained using ML estimation, optimizing

$$F_{ML} = \sum_{g=1}^{G} \frac{N_g}{N} F_{ML(g),} \tag{2}$$

where N_g is the sample size in study g and $N = N_1 + N_2 + \ldots + N_G$, and

$$F_{ML(g)} = \log|\Sigma_g| - \log|S_g| + \text{trace}\left(S_g \Sigma_g^{-1}\right) - p_g, \tag{3}$$

where S_g is the $p_g \times p_g$ sample variance–covariance matrix in the gth study. In applications of TSSEM, S_g is often not observed, but R_g is. In such cases, R_g can be substituted for S_g, and the estimated elements in D_g will be close to 1 (Cheung, 2015b).

A chi-square measure of fit for the model in Equation [1] is available by comparing its F_{ML} value with the F_{ML} value of a saturated model that is obtained by leaving out the selection matrix M_g and replacing P with P_g,

$$\Sigma_g = D_g \, P_g \, D_g, \tag{4}$$

with diag(P_g) = I for all g. The difference between the resulting F_{ML} values of the models in Equations [1] and [4], multiplied by $N - G$, has a chi-square distribution with degrees of freedom equal to the difference in numbers of free parameters. If the chi-square value of this likelihood ratio test is significant, then the hypothesis of homogeneity must be rejected.

If the correlation matrices are not homogeneous, it may not make sense to create a single pooled correlation matrix to test the structural equation model in Stage 2. Possible solutions are to create subgroups of studies with homogenous correlation matrices and to fit different Stage 2 models to each subgroup (Cheung and Chan, 2005b), or to allow study-level variation in the correlation coefficients by using a random effects approach (Cheung, 2014). In the present study, we assume that homogeneity holds, and we do not evaluate homogeneity or consider heterogeneous situations.

Stage 2: *Fitting structural equation models*

Cheung and Chan (2005a) proposed to use WLS estimation to fit structural equation models to the pooled correlation matrix P that is estimated in Stage 1. Fitting the Stage 1 model provides estimates of the population correlation coefficients in P as well as the asymptotic variances and covariances of these estimates, V. In Stage 2, hypothesized structural equation models are fitted to the Stage 1 estimates of P by minimizing the WLSs fit function (also known as the asymptotically distribution free fit function; Browne, 1984):

$$F_{WLS} = (r - \rho_{MODEL})^{T} V^{-1} (r - \rho_{MODEL}), \tag{5}$$

where r is a column vector with the estimates of the unique elements in P, indicated with r (instead of ρ-hat) because the estimates are treated as observed values in the Stage 2 model. ρ_{MODEL} is a column vector with the unique elements in the model implied correlation matrix (P_{MODEL}), and V^{-1} is the inversed matrix of asymptotic variances and covariances of the Stage 1 estimates, which is used as the weight matrix. For example, in order to fit a factor model with k factors, one would specify P_{MODEL} as

$$P_{MODEL} = \Lambda \, \Phi \, \Lambda^{T} + \Theta, \tag{6}$$

where Φ is a k by k covariance matrix of common factors, Θ is a q by q (diagonal) matrix with residual variances, and Λ is a q by k matrix with factor loadings. Minimizing the WLS function leads to correct parameter estimates
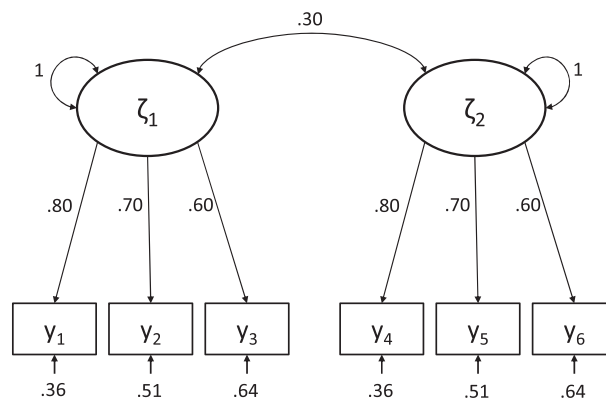

Figure 1. Population model used in data generation. Note: Fitting this model to data leads to eight degrees of freedom (6 * 7/2 = 21 observed statistics, 6 factor loadings + 6 residual variances + 1 factor correlation = 13 estimated parameters, df = 21 − 13 = 8). When fitting the model to a correlation matrix, the diagonal elements do not count as observed statistics, and the diagonal elements of the model implied matrix are constrained to be 1, leading to the same number of degrees of freedom.

with appropriate standard errors and a WLS-based chi-square test statistic T_{WLS} (Cheung and Chan, 2005a, Cheung, 2010).

Analyzing a correlation matrix as if it is a covariance matrix may lead to incorrect results (Cudeck, 1989). A solution to this problem is to constrain the diagonal of the model implied correlation matrix to unity during estimation ($\mathrm{diag}(\mathbf{P}_{MODEL}) = \mathrm{diag}(\mathbf{I})$).

Examples of fitting path models and factor models using TSSEM using the metaSEM-package can be found in Cheung (2015a, 2015b) and Jak (2015). In addition, we provide an example of fitting the 2-factor model from Figure 1 to a generated dataset based on 15 studies with a sample size of 200 each. The data and R script can be found in the Supporting Information.

3. Maximum likelihood meta-analytic structural equation modeling

As an alternative to using WLS for fitting a $\mathbf{P}_{MODEL}$ to the estimated $\mathbf{P}$ in a single-group analysis, we propose to retain the multigroup analysis and to use ML for fitting a common $\mathbf{P}_{MODEL}$ to the observed $\mathbf{R}_g$ matrices. ML MASEM consistently uses the individual studies' correlation matrices as input to the analysis. The structural equation model can be fitted to the observed $\mathbf{R}_g$ matrices by substituting $\mathbf{P}_{MODEL}$ for $\mathbf{P}$ in Equation [1],

$$\Sigma_g = \mathbf{D}_g\left(\mathbf{M}_g\,\mathbf{P}_{MODEL}\,\mathbf{M}_g{}^{\mathsf{T}}\right)\mathbf{D}_g. \tag{7}$$

Multigroup ML analysis of Equation [7] yields an ML chi-square of overall goodness-of-fit (T_{ML}) and correct standard errors, just as multigroup ML analysis of Equation [1]. Moreover, as $\mathbf{P}_{MODEL}$ in Equation [7] is a restriction of $\mathbf{P}$ in Equation [1], the difference between the associated chi-square values has a chi-square distribution itself with degrees of freedom equal to the difference in the numbers of free parameters in $\mathbf{P}$ and $\mathbf{P}_{MODEL}$.

For example, if we choose the factor model of Equation [6] as $\mathbf{P}_{MODEL}$, then the multigroup structural equation model is given by

$$\Sigma_g = \mathbf{D}_g\left(\mathbf{M}_g\left(\Lambda\,\Phi\,\Lambda^{\mathsf{T}} + \Theta\right)\mathbf{M}_g{}^{\mathsf{T}}\right)\mathbf{D}_g. \tag{8}$$

Just as with TSSEM, a constraint on the diagonal of the model implied correlation matrix is added ($\mathrm{diag}(\mathbf{P}_{MODEL}) = \mathrm{diag}(\mathbf{I})$), to ensure the correctness of the results when analyzing a correlation matrix. In summary, ML MASEM consists of three steps, involving three nested models:

- Step 0: The saturated model (Equation [4]),
- Step 1: The homogeneity model (Equation [1]), and
- Step 2: The structural equation model (Equation [7]).

Table 1 gives an overview of the three models and the associated tests. The difference in fit of the first two models provides a test of the homogeneity of correlation coefficients, and the difference in fit of the Steps 1 and 2 models provides a test of the structural equation model.

The ML MASEM models can be fitted using the OpenMx package (Boker *et al.*, 2011). In the Supporting Information, we provide annotated syntax to fit the two-factor model from Figure 1 to a simulated dataset with 15 studies with a sample size of 200 each.

3.1. A priori comparison of TSSEM and ML MASEM

The Step 1 model in ML MASEM is identical to the Stage 1 model in TSSEM. The difference between the two procedures lies in the way the structural equation model is fitted to the data. Where TSSEM fits the structural equation model to a single (estimated) pooled correlation matrix, ML MASEM fits the structural equation model to the (observed) correlation matrices of all studies in a multigroup model. As the Step 2 model in ML MASEM still is a multigroup model, you could say that the full three-step ML MASEM procedure stays within Stage 1 of TSSEM.

In ML MASEM, the Step 2 model is a special case of the Step 1 model, and the Step 1 model is a special case of the Step 0 model. As a result, one could evaluate the fit of the structural equation model with the chi-square based on the difference in fit of the Steps 0 and 2 models directly. A significant chi-square would then indicate

Table 1. An overview of models and associated tests in ML MASEM.

Model	Equation	Likelihood ratio test
Model 0 (saturated)	$\Sigma_g = \mathbf{D}_g\,\mathbf{P}_g\,\mathbf{D}_g^{\mathsf{T}}$	—
Model 1 (Step 1)	$\Sigma_g = \mathbf{D}_g\,(\mathbf{M}_g\,\mathbf{P}\,\mathbf{M}_g^{\mathsf{T}})\,\mathbf{D}_g^{\mathsf{T}}$	Homogeneity of correlations (Model 1 versus Model 0)
Model 2 (Step 2)	$\Sigma_g = \mathbf{D}_g\,(\mathbf{M}_g\,\mathbf{P}_{MODEL}\,\mathbf{M}_g^{\mathsf{T}})\,\mathbf{D}_g^{\mathsf{T}}$	Fit of structural equation model (Model 2 versus Model 1)

ML MASEM, maximum likelihood meta-analytic structural equation modeling.

heterogeneity of correlation matrices, or misfit of the structural equation model, or both. To separate these two sources of misfit, it is recommended to evaluate the homogeneity of the correlation matrices first (by comparing model fit in Steps 0 and 1) and to evaluate the fit of the structural equation model by comparing the fit of the Step 2 model with the fit of the Step 1 model. In the remainder of this article, we focus on the structural equation models, and all reported test statistics associated with ML MASEM are based on a comparison of the Step 1 model with the Step 2 model.

Maximum likelihood meta-analytic structural equation modeling may have some statistical advantages over TSSEM because the same estimation method is used for all models, and the Steps 1 and 2 models are nested. For this reason, it is interesting to compare the performance of ML MASEM and TSSEM in terms of estimation bias, false positive rates, and true positive rates. It is noted, however, that we already know that TSSEM yields parameter estimates, false positive rates, and true positive rates that are very close to the expected values (Cheung and Chan, 2005a). Still, we expect that the statistical performance of ML MASEM is at least as good as the performance of TSSEM.

4. Simulation study

We will use simulated data to investigate the performance of ML MASEM and compare it with TSSEM as proposed by Cheung and Chan (2005a). As the two methods involve identical homogeneity models, we will focus on the structural equation models and calculate estimation bias, false positive rates, and power using the structural equation models in both the multigroup ML estimation in ML MASEM and the single group WLS estimation in TSSEM. The design and population model of our simulation study are chosen similar to the design and population model in the Cheung and Chan (2005a) study.

As a result, we have numbers of studies equal to 5, 10, and 15, with 50, 100, 200, 500, and 1000 respondents each. The population model is a two-factor model with six variables, with parameter values for Λ, Φ, and Θ as shown in Figure 1. We generated continuous multivariate normal data using the MASS package in the R program (R Development Core Team, 2011), after which we computed the correlation matrices. As it is not realistic that all studies reported correlations between all variables, some variables were removed from some studies. Table 2 shows the patterns of missing variables in conditions with 5, 10, and 15 studies. These patterns are equal to the missing data patterns in Cheung and Chan (2005a). Conditions with more studies have more missing variables than conditions with fewer studies; the percentages of missing variables are equal to 20%, 33%, and 46% for conditions with 5, 10, and 15 studies, respectively. This seems to be realistic as researchers who want to include more studies in their meta-analysis will include more studies with missing variables. As the pattern is fixed a priori, the missingness may be considered missing completely at random (Graham *et al.*, 1996) and will not bias parameter estimates. We generated 500 datasets for each of the 15 conditions. The TSSEM and ML MASEM analyses were performed using the same datasets, to make sure that differences in the outcomes are due to differences between the methods and not due to differences in the data. With 500 datasets, the 95% prediction interval around the chosen α-level of 0.05 is [0.031–0.069].

Table 2. The patterns of missing data in conditions with 5, 10, and 15 studies.

	Conditions with 5 studies						Conditions with 10 studies						Conditions with 15 studies					
	Observed variable						Observed variable						Observed variable					
	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$
Study 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Study 2	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1
Study 3	1	0	1	1	1	0	1	0	1	1	1	0	1	0	1	1	1	0
Study 4	1	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	1
Study 5	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1
Study 6							0	0	0	1	1	1	0	0	0	1	1	1
Study 7							1	0	0	0	1	1	1	0	0	0	1	1
Study 8							1	1	0	0	0	1	1	1	0	0	0	1
Study 9							1	1	1	0	0	0	1	1	1	0	0	0
Study 10							0	1	1	1	1	0	0	0	1	1	1	0
Study 11													0	0	0	0	1	1
Study 12													1	0	0	0	0	0
Study 13													1	1	0	0	0	0
Study 14													0	1	1	0	0	0
Study 15													0	0	0	1	1	0

Note: 1 = variable present in study; 0 = variable absent in study.

4.1. Estimation bias

We investigated estimation bias by fitting the model of Figure 1 to the data with both ML MASEM and TSSEM. The factor variances are not estimated but fixed at unity for identification. The estimated parameters are as follows: six factor loadings, six residual variances, and one covariance between the two factors. The percentage of relative estimation bias is calculated as mean(estimate − population value) / population value * 100. According to Muthén *et al.* (1987), relative estimation bias less than 10% can be considered negligible. Because positive and negative bias will cancel each other out in the calculation of relative bias, we also evaluate absolute bias, calculated as mean(|estimate − population value|) / population value * 100. Differences in absolute bias in the estimation of factor loadings and factor covariances were tested through analysis of variance with "method" as a within factor with two levels (ML MASEM and TSSEM) and number of studies (three levels) and sample size of the studies (five levels) as between factors.

4.2. False positive rates

The distribution of T_{ML} (obtained from ML MASEM) and T_{WLS} (obtained from TSSEM) and false positive rates for the tests of overall goodness-of-fit were inspected by fitting the correct model (Figure 1) to the datasets (df = 8, $\alpha = 0.05$). The mean, standard deviation, and rejection rates of T_{ML} and T_{WLS}, as well as QQ-plots and Kolmogorov–Smirnov tests of the empirical versus theoretical chi-square distributions were used to evaluate the T_{ML} and T_{WLS} distributions. We also calculated false positive rates for the significance test of a single parameter estimate, by fitting a model in which Variable 4 has a redundant cross loading on the first factor. False positives for this single parameter estimate were determined in two ways. Firstly, on the basis of the likelihood ratio test, by checking the chi-square difference between models with and without the redundant parameter (df = 1, $\alpha = 0.05$). Secondly, on the basis of the 95% likelihood based confidence interval (Neale & Miller, 1997) of the estimate for the redundant parameter, by counting the number of times that the interval did not include zero. Differences in false positive rates associated with ML MASEM and TSSEM were tested using McNemar's test for dependent proportions, at a significance level of 0.05.

4.3. True positive rates (power)

To investigate the power to detect non-zero parameters, we generated data according to the population model given in Figure 1 but with an additional 0.10 loading of Variable 4 on the first factor. We fitted models with and without the cross loading to these datasets. Power was determined for the overall goodness-of-fit (by counting true positives of the chi-square difference test with df = 8, $\alpha = 0.05$) and for the test of the single additional 0.10 loading (chi-square difference test with df = 1, $\alpha = 0.05$). In addition, we counted the number of times that the 95% confidence interval of the additional parameter estimate did not include zero.

We compare the power of ML MASEM and TSSEM to each other and to the power that is expected on the basis of the non-centrality parameter that was estimated by fitting the covariance matrix implied by the Stage 2 model (Model 2 as in Figure 1) to the covariance matrix implied by the Stage 2 model with the cross loading (Model 2 with the additional loading) (Saris and Satorra (1993). Differences in true positive rates with ML MASEM and TSSEM were tested using McNemar's tests for dependent proportions, at a significance level of 0.05.

5. Results

5.1. Convergence

All results are based on only those datasets for which convergence was reached at all stages or steps of the modeling process of the particular method. The number of analyzed datasets varies between 493 and 500, but in most conditions, neither method showed any convergence problem (*cf.* the last two columns of Tables 3, 4, 6, and 7).

5.2. Estimation bias

Table 3 gives the percentages of relative estimation bias in a selection of parameters: the first three factor loadings and the covariance between the two common factors. Although ML MASEM generally shows less estimation bias than TSSEM, all bias is very small, ranging from 0.0% to (minus) 3.5%. Percentages of absolute bias in Table 4 are larger than the percentages of relative bias, ranging from 1.2% to 8.7% for the factor loadings, which is still well below the 10% to 15% considered satisfactory by Muthén *et al.* (1987). However, absolute bias in the factor covariance is more than 20% in the conditions with the smallest sample size. Absolute bias gets smaller with increasing sample size.

Although the differences in bias between ML MASEM and TSSEM are very small, analysis of variances on the absolute bias indicate that ML MASEM estimates are significantly less biased in conditions with smaller sample sizes ($N = 50$, $N = 100$). The results of the significance tests can be found in Table S1 in the Supporting Information.

Table 3. Percentages of relative estimation bias for selected parameters.

		λ_{11}		λ_{21}		λ_{31}		ϕ_{21}		#Converged	
		ML	WLS	ML	WLS	ML	WLS	ML	WLS	ML	WLS
5 studies	N = 50	0.448	1.228	0.331	1.794	−0.164	1.643	−1.208	1.671	498	497
	N = 100	0.765	1.217	−0.194	0.491	−0.277	0.593	−0.567	0.999	499	499
	N = 200	0.052	0.255	0.041	0.293	0.207	0.631	**−0.673**	**0.165**	500	500
	N = 500	0.008	0.092	0.015	0.143	0.043	0.222	0.178	0.511	500	500
	N = 1000	**−0.105**	**−0.057**	**−0.057**	**0.005**	0.257	0.333	**−0.134**	**0.055**	499	500
10 studies	N = 50	−0.058	0.555	0.990	1.764	0.469	1.806	1.431	3.499	500	500
	N = 100	−0.059	0.218	0.319	0.664	0.267	0.860	0.133	1.117	500	500
	N = 200	−0.060	0.086	0.143	0.305	0.175	0.446	0.636	1.153	500	500
	N = 500	0.161	0.216	**−0.157**	**−0.083**	**−0.064**	**0.050**	0.508	0.710	500	500
	N = 1000	0.131	0.162	**−0.037**	**−0.010**	−0.018	0.020	**−0.225**	**−0.105**	500	497
15 studies	N = 50	0.186	0.686	0.451	1.066	0.679	2.026	0.394	2.438	500	500
	N = 100	0.131	0.326	0.020	0.303	**−0.421**	**0.161**	0.665	1.641	500	499
	N = 200	0.383	0.478	0.000	0.143	0.018	0.268	0.855	1.303	499	500
	N = 500	0.076	0.131	0.093	0.168	0.214	0.341	0.002	0.126	493	500
	N = 1000	0.035	0.070	**−0.103**	**−0.089**	0.147	0.182	−0.157	−0.053	493	494
Overall		0.140	0.378	0.124	0.464	0.102	0.639	0.122	1.015		

ML, maximum likelihood; ML MASEM, maximum likelihood meta-analytic structural equation modeling; WLS, weighted least squares; TSSEM, two-stage structural equation modeling.

Note: ML = ML MASEM, WLS = TSSEM. Convergence indicates the number of datasets for which the analysis converged to a solution. Bold figures indicate cases in which TSSEM results were closer to zero than ML MASEM results. Relative bias is calculated per condition as mean(estimate − population value) / population value * 100.

Table 4. Percentages of absolute estimation bias for selected parameters.

		λ_{11}		λ_{21}		λ_{31}		ϕ_{21}		#Converged	
		ML	WLS	ML	WLS	ML	WLS	ML	WLS	ML	WLS
5 studies	$N=50$	6.982	7.133	7.566	7.759	8.429	8.721	22.385	22.861	498	497
	$N=100$	5.115	5.137	5.617	5.655	6.369	6.520	15.946	16.044	499	499
	$N=200$	3.494	3.475	3.651	3.656	4.516	4.565	11.086	11.175	500	500
	$N=500$	2.199	2.203	2.187	2.192	**2.657**	**2.655**	7.090	7.111	500	500
	$N=1000$	1.484	1.491	1.678	1.678	1.994	2.009	5.136	5.143	499	500
10 studies	$N=50$	6.164	6.237	6.220	6.357	7.029	7.408	20.354	20.899	500	500
	$N=100$	**3.980**	**3.973**	**4.495**	**4.492**	4.961	5.035	13.626	13.749	500	500
	$N=200$	**2.830**	**2.825**	**3.135**	**3.127**	3.557	3.586	9.796	9.843	500	500
	$N=500$	1.860	1.861	**1.925**	**1.916**	2.359	2.364	5.534	5.542	500	500
	$N=1000$	**1.237**	**1.236**	1.326	1.327	**1.643**	**1.636**	4.347	4.347	500	497
15 studies	$N=50$	5.738	5.704	6.147	6.289	7.431	7.732	18.855	19.185	500	500
	$N=100$	**4.034**	**3.976**	4.311	4.317	5.690	5.711	12.938	13.056	500	499
	$N=200$	2.707	2.738	3.042	3.046	3.714	3.736	9.583	9.663	499	500
	$N=500$	1.721	1.729	1.821	1.826	2.243	2.257	5.779	5.781	493	500
	$N=1000$	**1.211**	**1.209**	1.363	1.361	1.606	1.622	**4.142**	**4.138**	493	494
Overall		3.384	3.395	3.632	3.667	4.280	4.371	11.106	11.236		

ML, maximum likelihood; ML MASEM, maximum likelihood meta-analytic structural equation modeling; WLS, weighted least squares; TSSEM, two-stage structural equation modeling.
Note: ML = ML MASEM, WLS = TSSEM. Convergence indicates the number of datasets for which the analysis converged to a solution. Bold figures indicate cases in which TSSEM results were smaller than ML MASEM results. Absolute bias is per condition as: mean(|estimate − population value|) / population value * 100.

The main effect of number of studies and the main effect of sample size were significant for all parameters. For two of the parameters, their interaction effect was also significant. Inspection of the means per condition (Table 4) shows that absolute estimation bias decreases with larger sample sizes and with more studies, and that, for the first two factor loadings, the effect of sample size is stronger with lesser studies. There were significant main effects of method on all parameters except the first factor loading, indicating significantly less bias with ML MASEM than with TSSEM. For all parameters, there was a significant interaction effect of method and sample size on bias, indicating that the decrease in bias with increasing sample size is stronger with ML MASEM than with TSSEM.

Table 5. Means, standard deviations, and rejection rates of T_{ML} and T_{WLS} (df = 8, $\alpha = 0.05$).

		T_{ML} (ML MASEM)			T_{WLS} (TSSEM)		
		Mean	sd	Rejection rate	Mean	sd	Rejection rate
5 studies	$N=50$	8.283	3.964	0.066	8.383	4.045	0.068
	$N=100$	8.249	4.233	0.066	8.303	4.265	0.070
	$N=200$	**7.805**	**3.791**	0.048	**7.831**	**3.806**	0.048
	$N=500$	8.203	4.258	0.052	8.219	4.270	0.052
	$N=1000$	8.175	**4.160**	0.054	8.180	**4.159**	0.054
10 studies	$N=50$	8.464	4.348	0.064	8.640	4.520	0.074
	$N=100$	8.007	4.102	0.050	8.095	4.200	0.058
	$N=200$	8.014	3.988	0.050	8.053	4.030	0.052
	$N=500$	8.185	4.093	0.062	8.199	4.108	0.062
	$N=1000$	8.009	3.937	0.042	8.050	3.934	0.042
15 studies	$N=50$	8.524	4.130	0.078*	8.766	4.372	0.092*
	$N=100$	8.076	4.198	0.062	8.151	4.285	0.066
	$N=200$	**7.924**	**3.862**	**0.044**	**7.973**	**3.921**	**0.046**
	$N=500$	**7.785**	**3.797**	**0.037**	**7.855**	**3.844**	**0.042**
	$N=1000$	8.148	4.014	0.055	8.231	4.037	0.059

ML MASEM, maximum likelihood meta-analytic structural equation modeling; TSSEM, two-stage structural equation modeling.
Note: Bold figures indicate cases in which TSSEM results were closer to the expected mean (8.00), standard deviation (4.00), or rejection rate (0.05).
*The rejection rates for ML MASEM and TSSEM are significantly different at $\alpha = 0.05$.

Table 6. False positives rates for redundant model parameter.

		LRT ($\alpha = 0.05$)		95% CI		#Converged	
		ML MASEM	TSSEM	ML MASEM	TSSEM	ML MASEM	TSSEM
5 studies	$N = 50$	0.057	0.063	0.055	0.063	493	492
	$N = 100$	**0.050**	**0.044**	**0.050**	**0.044**	498	498
	$N = 200$	0.050	0.052	0.050	0.052	500	500
	$N = 500$	0.046	0.048	0.046	0.048	500	500
	$N = 1000$	0.056	0.060	0.054	0.060	499	499
10 studies	$N = 50$	0.040	0.042	0.040	0.042	497	496
	$N = 100$	0.044	0.048	0.044	0.048	499	500
	$N = 200$	0.046	0.046	0.046	0.046	499	499
	$N = 500$	**0.068**	**0.066**	0.066	0.066	500	499
	$N = 1000$	0.052	0.052	0.052	0.052	500	496
15 studies	$N = 50$	0.048	0.050	0.048	0.050	498	499
	$N = 100$	0.058	0.058	0.058	0.056	497	497
	$N = 200$	0.068	0.068	**0.068**	**0.066**	499	499
	$N = 500$	**0.055**	**0.054**	**0.057**	**0.054**	493	500
	$N = 1000$	**0.047**	**0.046**	**0.047**	**0.046**	493	496

ML MASEM, maximum likelihood meta-analytic structural equation modeling; TSSEM, two-stage structural equation modeling; CI, confidence interval; WLS, weighted least squares; ML, maximum likelihood; LRT, likelihood ratio test.
Note: Bold figures indicate cases in which WLS has better (lower) false positive rates than ML. Proportions are calculated based on converged solutions only. Rejection rates do not differ significantly between ML and WLS in any of the conditions at $\alpha = 0.05$.

Overall, ML MASEM is doing significantly better than TSSEM, although the differences are so small that they may be negligible in practice. In the Supporting Information (Tables S2 and S3), we added the average width and the coverage rates of the 95% likelihood-based confidence intervals for the selection of parameters. The estimates and coverage rates of the confidence intervals were very similar for ML MASEM and TSSEM.

5.3. Distributions of the test statistics

Table 5 gives the means and standard deviations of the test statistics obtained with ML MASEM (T_{ML}) and TSSEM (T_{WLS}). In all conditions, both test statistics approximate the chi-square distribution well, as the mean and standard

Table 7. Theoretical (Exp.) and empirical (ML and WLS) power of the overall test, the likelihood ratio test, and the likelihood-based confidence intervals to detect a cross loading of 0.10 ($\alpha = 0.05$).

	N	Overall test			LRT			95% CI		#Converged	
		Exp.	ML	WLS	Exp.	ML	WLS	ML	WLS	ML	WLS
5 studies	50	0.102	0.114	0.127	0.216	0.236	0.254	0.242	0.254	496	495
	100	0.170	0.170	0.160	**0.384**	**0.377**	**0.388**	**0.377**	**0.390**	499	499
	200	0.331	0.317	0.311	**0.656**	**0.643**	**0.647**	**0.643**	**0.647**	499	499
	500	0.764	0.736	0.724	0.962	0.962	0.960	0.962	0.960	497	500
	1000	0.983	0.984	0.980	1.000	1.000	1.000	1.000	1.000	499	497
10 studies	50	0.118	0.112	0.133	0.259	0.264	0.283	0.266	0.281	498	498
	100	0.208	0.210	0.222	0.462	0.452*	0.471*	0.452*	0.471*	500	499
	200	**0.416**	**0.396**	**0.400**	0.753	0.718*	0.730*	0.718*	0.730*	500	500
	500	0.868	0.830	0.828	0.987	0.986	0.986	0.986	0.986	500	500
	1000	0.997	0.998	0.998	1.000	1.000	1.000	1.000	1.000	500	493
15 studies	50	0.113	0.112	0.130	**0.246**	**0.222**	**0.234**	**0.222**	**0.232**	495	500
	100	0.196	0.244	0.254	0.440	0.436	0.432	0.434	0.436	500	498
	200	0.391	0.400	0.400	**0.727**	0.726	0.724	**0.726**	**0.730**	500	497
	500	0.842	0.852	0.852	0.982	0.974	0.974	0.974	0.974	494	500
	1000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	499	493

ML, maximum likelihood; WLS, weighted least squares; ML MASEM, maximum likelihood meta-analytic structural equation modeling; TSSEM, two-stage structural equation modeling; LRT, likelihood ratio test.
Note: ML = ML MASEM, WLS = TSSEM. Bold figures indicate cases in which TSSEM power is closer to the expected power that is based on the non-centrality parameter than ML MASEM.
*The tests rejection rates for ML MASEM and TSSEM within the condition are significantly different at $\alpha = 0.05$.

deviations are very close to their expected values (mean = 8, sd = 4), with the T_{ML} values generally slightly closer than the T_{WLS} values. Table 4 also shows the false positive rates, that is, the rates of rejecting the correct model. The false positive rates appear to converge to the expected value (0.05) when the sample size increases, which is in accordance with the findings of Cheung and Chan (2005a). The results are very similar for ML MASEM and TSSEM, with the only statistically significant difference found in the condition with 15 studies with sample sizes of 50. In the Supporting Information, we provide QQ-plots and the results of Kolmogorov–Smirnov tests comparing the distributions of T_{ML} and T_{WLS} with the theoretical distribution and with each other. The plots and tests indicate that there are no significant differences between the distributions of T_{ML} and T_{WLS} in any of the conditions. In the condition with 15 studies of size 50, both T_{ML} and T_{WLS} differed significantly from the theoretical distribution ($p = 0.03$ for T_{ML}, $p < 0.01$ for T_{WLS}). In the condition with 10 studies of size 50, only T_{WLS} differed significantly ($p = 0.04$).

5.4. False positive rates of redundant parameter

The false positive rates of the redundant cross loading are shown in Table 6. These were very close to the expected 0.05 in all conditions and were not found to differ significantly between ML MASEM and TSSEM.

5.5. True positive rates (power)

Table 7 gives the true positive rates or power to reject the model without a cross loading for the overall goodness-of-fit test (df = 8), for single parameter test (df = 1), and for the 95% confidence interval. As expected, the power increases with sample size. The number of studies does not affect power that much, probably because the numbers of studies do not vary much (5, 10, and 15) and the number of missing variables increases with the number of studies. Both methods reach acceptable power levels with sample sizes of 500 or larger. ML MASEM generally shows power results that are somewhat closer to the expected power than TSSEM, but the TSSEM often yields slightly more true positives. In the two conditions with ten studies and sample sizes of 100 or 200, the rejection rate for TSSEM was found to be significantly higher than the rejection rate for ML MASEM. The power in the conditions with 15 studies is somewhat lower than in the conditions with ten studies. This is caused by the simulation design with larger numbers of missing values in conditions with 15 studies. Specifically, none of the five studies that were added in the 15-study condition provided correlations for Variable 4 with the three indicators of the first factor (Variables 1 to 3), because either Variable 4 was missing or Variables 1 to 3 were missing. When we evaluated the power of the test on a cross loading of Variable 4 on the first factor, the datasets in the conditions with 15 studies therefore did not contain any additional information regarding this cross loading.

6. Discussion

The simulation study showed that ML MASEM and TSSEM generally yield very similar results, although ML MASEM parameter estimates are significantly less biased when sample sizes are small. Test statistics, confidence intervals, false positive rates, and true positive rates do not differ between the two methods.

6.1. Advantages and disadvantages of ML MASEM

Although the differences were not large, ML MASEM leads to significantly less bias in parameter estimates than TSSEM. Besides differences in average estimation bias, there are some fundamental and practical differences, which may guide a researcher's choice between the two methods. A theoretical advantage of ML MASEM is that the same estimation method is used for all models involved, so that the structural equation model can be evaluated as a special case of the homogeneity model.

Maximum likelihood meta-analytic structural equation modeling is also more flexible in the application of equality constraints across studies in the structural equation model: Some parameters in the structural equation model could be set equal across a subset of studies, another parameter could be set equal across another subset of studies and yet other parameters could be freely estimated in all studies, provided that the (groups of) studies contain enough identification information. Free estimation of model-implied correlation coefficients in specific studies is always possible, but if one wants to free a specific parameter for a group of studies, then these studies must of course contain sufficient information to identify this parameter.

Maximum likelihood meta-analytic structural equation modeling can be conducted with OpenMx, but specifying the OpenMx script is laborious. It would be useful to automate the writing of the OpenMx script. For now, we have made an example OpenMx script available as Supporting Information, which can be modified to fit new examples.

The example script also shows how the root mean square error of approximation (RMSEA) and comparative fit index (CFI) of the structural equation model can be calculated using the output of the different models. The RMSEA is calculated using the chi-square difference between Models 2 and 1, that is, the chi-square value that represents the misfit of the structural model. The CFI is calculated using this chi-square and the chi-square

difference obtained by fitting an independence model as Model 2. We are not aware of earlier work showing the correctness of the CFI and RMSEA obtained in this way. However, as the chi-square values for the structural model are found to be very similar for ML MASEM and TSSEM in our simulation study, the calculated fit indices will also be very similar.

6.2. Advantages and disadvantages of TSSEM

Two-stage structural equation modeling has practical advantages over ML MASEM. In TSSEM, the structural equation model is not a multigroup model, so that estimation convergence is much faster than in ML MASEM. The necessity to calculate a weight matrix (the inverse of the matrix of asymptotic variances and covariances of the pooled correlation coefficients) may count as a disadvantage of TSSEM, but fortunately the readily available R package metaSEM takes this burden off of the user. As a result, TSSEM may actually be easier to use than ML MASEM.

Another possible disadvantage of TSSEM is that the pooled correlation matrix that is estimated in the multigroup SEM in Stage 1 is taken as an observed matrix in the single group SEM in Stage 2. In ML MASEM, misfit in the structural model may be partly alleviated by adjusting the estimates of the pooled correlations, which is not possible in TSSEM, where the pooled correlation matrix of Stage 1 is not part of the model. This could lead to larger rejection rates with TSSEM. Still, in our study, TSSEM did not show more false positives than ML MASEM, but it did yield slightly more true positives. Notably, true positive rates for both methods were generally somewhat higher than expected.

6.3. Handling heterogeneity with TSSEM and ML MASEM

In applications of MASEM, correlation matrices will often not be homogenous, rendering the fixed effects Stage 2 models inappropriate. One of the options to handle heterogeneity is to create subgroups of studies that can be considered homogeneous with respect to the correlation coefficients. These subgroups can be created based on study-level variables such as type of respondents in the study, experimental versus non-experimental studies, and type of measurement instruments used. If no interesting study-level variables are available, it may be an option to use a cluster analytic approach (Cheung and Chan, 2005b).

Another alternative is to use random effects modeling, which does not assume homogeneity of effect sizes across studies. Cheung (2013, 2014) shows how the Stage 1 model can be estimated using the random effects approach, after which the study-level variance is accounted for, and the Stage 2 model can be fitted to the pooled correlation matrix using the asymptotic variances and covariances from Stage 1 as the weight matrix in WLS estimation.

Maximum likelihood meta-analytic structural equation modeling has not yet been extended to include random effects. It would be useful to be able to estimate study-level variance for the Stage 2 parameters directly. As mentioned earlier, one way in which ML MASEM can handle heterogeneity is to freely estimate one or more of the parameters across studies.

6.4. Future research

We compared the performance of ML MASEM with TSSEM, but not to other methods, as in earlier research TSSEM was found to have the best statistical properties of the available MASEM methods (Cheung and Chan, 2005a). However, TSSEM and ML MASEM are not yet compared with newer versions of the GLS method (Hafdahl, 2007), which could be the focus of future research.

As mentioned before, the calculation of fit indices like the RMSEA and CFI within ML MASEM is new. Further research is needed to evaluate the appropriateness of these fit indices.

The simulation study we performed was small and only served to investigate whether ML MASEM is a viable alternative to TSSEM. The results obtained in our study cannot be generalized to the broad range of realistic conditions one could imagine. Although we found no differences between TSSEM and ML MASEM in the evaluated conditions, there may still be situations in which one of the methods outperforms the other. It could, for example, be that with larger and more complex models, in combination with small numbers of studies and small sample sizes, the estimated weight matrix in TSSEM becomes unstable, leading to a better performance with ML MASEM. Future studies may examine the performance of the methods with different numbers of studies, more variables, more complex models, varying sample sizes, and with more variation in numbers and patterns of missing variables.

6.5. Conclusion

Our simulation study showed that using ML MASEM leads to almost identical results as using TSSEM. ML MASEM estimates are less biased than TSSEM estimates, but the differences are very small. ML MASEM is a viable alternative method to TSSEM.

Research Synthesis Methods

Acknowledgements

References

Becker BJ 1992. Using results from replicated studies to estimate linear models. *Journal of Educational Statistics* **17**: 341–362.

Becker BJ 1995. Corrections to "using results from replicated studies to estimate linear models". *Journal of Educational and Behavioral Statistics* **20**: 100–102.

Becker BJ 2009. Model-based meta-analysis. In Cooper H, Hedges LV, Valentine JC (eds.). The Handbook of Research Synthesis and Meta-analysis (pp. 377–395). 2nd edn. New York, NY: Sage.

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J 2011. OpenMx: an open source extended structural equation modeling framework. *Psychometrika* **76**: 306–317.

Browne MW 1984. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* **37**: 62–83.

Cheung MW-L 2010. Fixed-effects meta-analyses as multiple-group structural equation models. *Structural Equation Modeling* **17**: 481–509.

Cheung MW-L 2013. Multivariate meta-analysis as structural equation models. *Structural Equation Modeling* **20**: 429–454.

Cheung MW-L 2014. Fixed-and random-effects meta-analytic structural equation modeling: examples and analyses in R. *Behavior Research Methods* **46**(1): 29–40.

Cheung MW-L 2015a. MetaSEM: an R package for meta-analysis using structural equation modeling. *Frontiers in Psychology* **5**: 1521. DOI:10.3389/fpsyg.2014.01521.

Cheung MW-L 2015b. Meta-analysis: A Structural Equation Modeling Approach. Chichester, UK: Wiley.

Cheung MW–L, Chan W 2005a. Meta-analytic structural equation modeling: a two-stage approach. *Psychological Methods* **10**: 40–64.

Cheung MW-L, Chan W 2005b. Classifying correlation matrices into relatively homogeneous subgroups: a cluster analytic approach. *Educational and Psychological Measurement* **65**: 954–979.

Cheung MW-L, Chan W 2009. A two-stage approach to synthesizing covariance matrices in meta-analytic structural equation modeling. *Structural Equation Modeling* **16**: 28–53.

Cudeck R 1989. Analysis of correlation matrices using covariance structure models. *Psychological Bulletin* **105**(2): 317–327.

Graham JW, Hofer SM, MacKinnon DP 1996. Maximizing the usefulness of data obtained with planned missing value patterns: an application of maximum likelihood procedures. *Multivariate Behavioral Research* **31**(2): 197–218.

Hafdahl AR 2007. Combining correlation matrices: simulation analysis of improved fixed-effects methods. *Journal of Educational and Behavioral Statistics* **32**: 180–205.

Hafdahl AR 2008. Combining heterogeneous correlation matrices: simulation analysis of fixed-effects methods. *Journal of Educational and Behavioral Statistics* **33**: 507–533.

Hedges LV, Olkin I 1985. Statistical Methods for Meta-analysis. Orlando, FL: Academic Press.

Hunter JE, Schmidt FL (Eds) 2004. Methods of Meta-analysis: Correcting Error and Bias in Research Findings. Newbury Park, CA: Sage.

Jak S 2015. Meta-analytic structural equation modeling. Springer International Publishing.

Muthén B, Kaplan D, Hollis M 1987. On structural equation modeling with data that are not missing completely at random. *Psychometrika* **52**: 431–462.

Neale MC, Miller MB 1997. The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics* **27**: 113–120.

R Development Core Team 2011. R: A Language and Environment for Statistical Computing. Computer software. Vienna: Austria.

Saris WE, Satorra A 1993. Power evaluations in structural equation models. In Bollen KA, Long JS (eds.). *Testing structural equation models* (pp. 181–204). Newbury Park, CA: Sage.

Viswesvaran C, Ones D 1995. Theory testing: combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology* **48**: 865–885.

Zhang Y. 2011. Meta-analytic structural equation modeling (MASEM): comparison of methods. Electronic Theses, Treatises and Dissertations. Paper 534.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.