Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R

Mike W.-L. Cheung

Abstract Meta-analytic structural equation modeling (MASEM) combines the ideas of meta-analysis and structural equation modeling for the purpose of synthesizing correlation or covariance matrices and fitting structural equation models on the pooled correlation or covariance matrix. Cheung and Chan (Psychological Methods 10:40–64, 2005b, Structural Equation Modeling 16:28–53, 2009) proposed a two-stage structural equation modeling (TSSEM) approach to conducting MASEM that was based on a fixed-effects model by assuming that all studies have the same population correlation or covariance matrices. The main objective of this article is to extend the TSSEM approach to a random-effects model by the inclusion of study-specific random effects. Another objective is to demonstrate the procedures with two examples using the `metaSEM` package implemented in the R statistical environment. Issues related to and future directions for MASEM are discussed.

Keywords Structural equation modeling · Meta-analysis · Meta-analytic structural equation modeling · Random-effects model

Structural equation modeling (SEM) is a popular statistical technique in the social, behavioral, and educational sciences. Part of its popularity is due to its flexibility in testing proposed or hypothesized models. The proposed models can be path models, confirmatory factor-analytic models, or general structural equation models that include latent and observed variables. Complex theories can be formulated as testable hypotheses in the form of structural equation models. The proposed models can be empirically tested by the use of a likelihood ratio (LR) statistic and various goodness-of-fit indices, and the estimated standard errors (*SE*s) can be used to test the significance of the parameter estimates.

Although SEM is very powerful for testing theories and hypothesized models, several unresolved issues still remain. One of these is that most researchers only compare a small number of models (MacCallum & Austin, 2000). Overall, researchers rarely consider alternative models. Also, the statistical power of rejecting incorrect models in SEM may not be high enough when the sample sizes are small. Due to this confirmation bias, the reported models in the literature may not be the correct (or the best) models. When a pool of empirical studies use similar variables, researchers may want to compare and synthesize these findings. It is difficult to address this issue in the current SEM framework.

In the behavioral sciences, meta-analysis is widely used as a statistical approach to synthesize research findings (e.g., Cooper, Hedges, & Valentine, 2009; Hedges & Olkin, 1985; Hunter & Schmidt, 2004). It has been successfully applied in various disciplines, including psychology, education, management, and the medical sciences. The limitations of synthesizing findings in SEM can be effectively addressed by combining SEM and meta-analysis—a technique widely known as *meta-analytic structural equation modeling* (MASEM; Cheung & Chan, 2005b).

MASEM combines the ideas of SEM and meta-analysis by pooling studies in order to draw general conclusions (Landis, 2013; Viswesvaran & Ones, 1995). MASEM has frequently been used to synthesize studies using SEM in the literature—for example, by testing the structural equivalence of the social axiom across 40 societies (Cheung, Leung, & Au, 2006), synthesizing 300 correlation matrices on Schwartz's theory of human values (Steinmetz, Baeuerle, & Isidor, 2012), and testing a mediation model from competition to performance, with performance-approach and performance-avoidance goals as specific mediators, across 474 studies (Murayama & Elliot, 2012).

M. W.-L. Cheung (✉)
Department of Psychology, Faculty of Arts and Social Sciences,
National University of Singapore, Block AS4, Level 2, 9 Arts Link,
Singapore 117570 Singapore
e-mail: mikewlcheung@nus.edu.sg

Different terms have been used for the techniques of combining SEM and meta-analysis in the literature for the purpose of fitting structural equation models to a pool of correlation or covariance matrices—for instance, *meta-analytic path analysis* (Colquitt, LePine, & Noe, 2000), *meta-analysis of factor analysis* (G. Becker, 1996), *SEM of a meta-analytic correlation matrix* (Conway, 1999), *meta-analytical structural equation analysis* (Hom, Caranikas-Walker, Prussia, & Griffeth, 1992), *path analysis of meta-analytically derived correlation matrices* (Eby, Freeman, Rush, & Lance, 1999), *path analysis on the basis of meta-analytic findings* (Tett & Meyer, 1993), and *model-based meta-analysis* (B. J. Becker, 2009). In this article, I will use the generic term *MASEM* to describe this class of techniques.

Cheung and Chan (2005b, 2009) proposed a two-stage SEM (TSSEM) approach to conducting fixed-effects MASEM. Several distinct features define this approach: First, multiple-group SEM is used to pool correlation or covariance matrices in the first stage of the analysis. Therefore, goodness-of-fit indices in SEM are available to test the homogeneity of the correlation or covariance matrices. Then, weighted least squares (WLS) are used to weigh the precision of the pooled correlation or covariance matrix in fitting structural models in the second stage of analysis. Thus, this approach allows different elements of the pooled correlation matrix to be weighted differently in fitting the structural models. Appropriate goodness-of-fit indices and *SE*s can be obtained.

The main objective of this article is to extend the fixed-effects TSSEM to a random-effects model. The second objective is to demonstrate how the proposed methods can be analyzed using the `metaSEM` package (Cheung, 2013b), which is based on the `OpenMx` package (Boker et al., 2011) implemented in the R statistical environment (R Development Core Team, 2013). The rest of the article is organized as follows. In the next section, the fixed-effects TSSEM is first reviewed. Next, the model is extended to the random-effects model. Two sample data sets from Digman (1997) and B. J. Becker and Schram (1994) are used to illustrate the procedures. Future directions in this line of research are discussed in the final section. Although I focus mainly on the analysis of correlation matrices, the techniques are readily applicable to the analysis of covariance matrices.

Two-stage structural equation modeling

Two stages are used when conducting a TSSEM. In the first stage, correlation matrices are pooled together. If a fixed-effects model is used, the homogeneity of the correlation matrices is tested. This test is known as the Q statistic in meta-analysis. If a random-effects model is used, the degree of heterogeneity of the correlation elements can be qualified by I^2. In the second stage, the pooled correlation matrix is

used to fit the proposed structural models. In this section, I will first present the fixed-effects models, and then extend them to random-effects models.

Fixed-effects models

The two classes of models in meta-analysis are *fixed-effects* and *random-effects* models (see, e.g., B. J. Becker, 1992, 1995; Hedges & Vevea, 1998; Schmidt, Oh, & Hayes, 2009). Fixed-effects models are used for conditional inferences based on the selected studies, and are intended to draw conclusions on the studies included in the meta-analysis. These models usually, but not always, assume that all studies share common effect sizes (cf. Bonett, 2009; Shuster, 2010, for different views). In the context of MASEM, applying a fixed-effects model assumes that the population correlation matrices are the same for all studies.

Stage 1 analysis Under the assumption of homogeneity of correlation matrices, the correlation or covariance matrix in the ith study can be decomposed as a correlation matrix ($\mathbf{P}_i$) and a diagonal matrix of standard deviations ($\mathbf{D}_i$) by

$$\Sigma_i(\theta) = \mathbf{D}_i \mathbf{P}_i \mathbf{D}_i \tag{1}$$

As is discussed by Cheung and Chan (2005b), $\mathbf{D}_i$ is required in order to correctly apply the statistical theory of analysis of the covariance matrix to the analysis of the correlation matrix. To obtain a pooled correlation matrix under the assumption of homogeneity, we may impose the constraints of $\mathbf{P}_{\text{Fixed}} = \mathbf{P}_1 = \mathbf{P}_2 = ... = \mathbf{P}_k$, where $\mathbf{D}_i$ may vary across studies. When there are incomplete correlation coefficients, the missing correlations are excluded from the constraints.

This approach has several advantages. First, missing or incomplete correlation elements can easily be handled by the maximum likelihood estimation (MLE) method (Allison, 1987; Muthén, Kaplan, & Hollis, 1987). When the missing mechanism is missing completely at random (MCAR) or missing at random (MAR), the parameter estimates using MLE are unbiased and efficient. Second, as is shown in Eq. 1, the Stage 1 analysis of the TSSEM approach does not need to estimate the sampling covariance $\mathbf{V}_i$ of the correlation coefficients $\mathbf{P}_i$. Thus, the TSSEM approach is very stable and accurate. In contrast, it is necessary to estimate the sampling covariance matrix of the correlation coefficients under the conventional generalized least squares (GLS) approach (e.g., B. J. Becker, 1992). For example, $\mathbf{V}_i$ is a 10×10 matrix if $\mathbf{P}_i$ is a 5×5 correlation matrix. $\mathbf{V}_i$ is treated as fixed in the GLS approach. Treating $\mathbf{V}_i$ as known values may affect the accuracy of the estimation, especially when the sample sizes are small. Thus, the GLS approach does not perform well empirically (see the simulation results in Cheung & Chan, 2005b,

2009). Third, the asymptotic covariance matrix $\widehat{\mathbf{V}}_{Fixed}$ of $\widehat{\mathbf{P}}_{Fixed}$ that indicates the precision of the estimates is routinely available after the analysis. The *SE* may be used to test the significance or to construct the approximate confidence intervals (CIs) of the pooled correlation or covariance matrices. Fourth, besides testing the assumption of the homogeneity of correlation matrices with an LR statistic, many goodness-of-fit indices, such as RMSEA and SRMR, may also be used to test the close or approximate fit of the homogeneity of the correlation matrices.

If the scales of the measurements are comparable across studies, researchers may want to synthesize the covariance matrices rather than the correlation matrices (e.g., Beretvas & Furlow, 2006). By testing the covariance matrices, researchers may test the measurement properties of the models across studies. Cheung and Chan (2009) showed that the TSSEM approach can be easily extended to synthesize covariance matrices by

$$\boldsymbol{\Sigma}_i(\theta) = \mathbf{S}_i, \tag{2}$$

where $\mathbf{S}_i$ is the sample covariance matrix in the *i*th study. To obtain a pooled covariance matrix, we may impose the following equality constraint: $\mathbf{S}_{Fixed} = \mathbf{S}_1 = \mathbf{S}_2 = ... = \mathbf{S}_k$. If there are incomplete covariances, they are excluded from the equality constraints.

Stage 2 analysis When dealing with correlation or covariance matrices, it is easier to transform the correlation or covariance matrices into vectors. Suppose that we have a symmetric matrix.

$$\mathbf{X} = \begin{bmatrix} 1 & & \\ 2 & 4 & \\ 3 & 5 & 6 \end{bmatrix}.$$

We may define $\mathbf{r} = \text{vechs}(\mathbf{X}) = [2\ 3\ 5]^T$ and $\mathbf{s} = \text{vech}(\mathbf{X}) = [1\ 2\ 3\ 4\ 5\ 6]^T$, which stack the lower triangle elements of the matrix by column major. Since the diagonal elements in a correlation matrix are always 1, elements in the diagonals are always excluded.

After the Stage 1 analysis, on the basis of a fixed-effects model, a vector of the pooled correlation matrix $\widehat{\boldsymbol{\rho}}_{Fixed} = \text{vechs}\left(\widehat{\mathbf{P}}_{Fixed}\right)$ and its asymptotic sampling covariance matrix $\widehat{\mathbf{V}}_{Fixed}$ are estimated. Many researchers treat $\widehat{\boldsymbol{\rho}}_{Fixed}$ as if it were an observed covariance matrix and use it to fit structural equation models. Cheung and Chan (2005b, 2009) identified several issues with this practice. One of them is that the pooled correlation matrix is analyzed as if it were a covariance matrix, so the LR statistics and the *SE* of the parameter estimates may be incorrect (Cudeck, 1989). Another issue is that the sampling variability of the estimated correlation

matrix $\widehat{\mathbf{V}}_{Fixed}$ has not been properly taken into account when the pooled correlation matrix is treated as the observed correlation matrix.

As it is likely that correlations will be missing, elements of the pooled correlation matrix may be based on different sample sizes. The third issue is that different researchers use different methods to determine the sample size in fitting the structural equation models. These include the arithmetic or geometric means for medium and the largest values of the sample sizes. However, these suggestions are all ad hoc and without strong statistical support. More importantly, using different sample sizes may lead to different *SE*s, test statistics, and goodness-of-fit indices.

To solve the aforementioned problems in the Stage 2 analysis, Cheung and Chan (2005b, 2009) proposed using WLS to fit the structural equation models. Suppose that the proposed structural model on the population correlation vector in the Stage 2 analysis is $\boldsymbol{\rho}(\boldsymbol{\gamma})$—that is, the population correlation vector $\boldsymbol{\rho}$ is a function of the unknown parameters $\boldsymbol{\gamma}$. The discrepancy function $F(\widehat{\boldsymbol{\gamma}})$ is

$$F\left(\widehat{\boldsymbol{\gamma}}\right) = \left[\boldsymbol{\rho}_{\text{Fixed}} - \boldsymbol{\rho}\left(\widehat{\boldsymbol{\gamma}}\right)\right]^T \mathbf{V}^{-1}_{\text{Fixed}} \left[\boldsymbol{\rho}_{\text{Fixed}} - \boldsymbol{\rho}\left(\widehat{\boldsymbol{\gamma}}\right)\right]. \tag{3}$$

$\boldsymbol{\rho}_{\text{Fixed}}$ and $\mathbf{V}_{\text{Fixed}}$ are taken from the Stage 1 analysis. Since they are treated as observed or fixed values in the Stage 2 analysis, no hats appear with either $\boldsymbol{\rho}_{\text{Fixed}}$ or $\mathbf{V}_{\text{Fixed}}$ in Eq. 3. This approach is generally known as WLS or the asymptotically distribution-free method (Browne, 1984).

When data are missing in the Stage 1 analysis, the estimated correlation coefficients with missing data will be less precise, with larger sampling variances and covariances. The logic of WLS estimation is to weight the correlation elements by the inverse of its sampling covariance matrix. Different weights are put into different elements in the pooled correlation matrix, depending on their precisions. Cheung (2010) shows that this WLS estimation function in SEM is related to the fixed-effects meta-analysis. This approach also automatically handles the sample size issue: Since the discrepancy function weights the correlation elements by their precision, the choice of sample size in the Stage 2 analysis does not affect the chi-square test and the estimated *SE*s computed. By using the WLS estimation function, parameter estimates with appropriate *SE*s, test statistics, and goodness-of-fit indices can be obtained in the Stage 2 analysis.

One potential criticism of using WLS as the estimation method is that large samples (at least 1,000) are usually required. Many simulation studies have shown that WLS performs badly in small samples (e.g., Curran, West, & Finch, 1996). Since MASEM is usually based on many studies, the resultant sample sizes are reasonably large. On the basis of simulations conducted by Cheung and Chan (2005b, 2009), the Stage 2 analysis using WLS estimation

was found to perform very well for ten studies with $N = 100$ per study. Thus, a total sample of 1,000 should be reasonable enough for the WLS estimation method.

Extension to categorical moderators When studies are heterogeneous, it is questionable to pool them with a fixed-effects model. One approach would be to group the studies into relatively homogeneous subgroups (Cheung & Chan, 2005a). If categorical study characteristics are present, such as intervention types in clinical studies and conditions in experiments, we can classify the studies into groups and conduct a fixed-effects TSSEM for each group. Instead of having one pooled correlation matrix, we might have several pooled correlation matrices, and proposed models could be fitted against the pooled correlation matrices.

Random-effects models

Random-effects models assume that the population correlation matrices may vary across studies, by assuming that the selected studies are random samples from a larger population. Suppose

that the model at the population is $\rho_{\text{Random}} = \text{vechs}[\mathbf{P}(\gamma)]$. Because of the random effects, each study has its own study-specific random effects,

$$\rho_i = \rho_{\text{Random}} + \mathbf{u}_i, \tag{4}$$

where ρ_i and $\mathbf{u}_i$ are the population correlation vector and the study-specific random effects in the ith study, respectively.

Stage 1 analysis A random-effects model for the correlation vectors $\mathbf{r}_i = \text{vechs}(\mathbf{R}_i)$ in the ith correlation matrix $\mathbf{R}_i$ is

$$\mathbf{r}_i = \rho_{\text{Random}} + \mathbf{u}_i + \mathbf{e}_i, \tag{5}$$

where $\text{Cov}(\mathbf{u}_i) = \mathbf{T}^2$ is the variance component of the study-specific random effects, and $\text{Cov}(\mathbf{e}_i) = \mathbf{V}_i$ is the known sampling covariance matrix in the ith study. $\mathbf{V}_i$ can be estimated using the model in Eq. 1 (see Cheung & Chan, 2004).

To fit the model above, B. J. Becker (1992) proposed a method-of-moments approach to obtain the parameter estimates and their *SE*s. Cheung (in press-b) showed how SEM can be used to conduct multivariate meta-analysis with the MLE method. The log-likelihood of the ith study under a random-effects meta-analysis is

$$\log l\left(\rho_{\text{Random}}, \mathbf{T}^2\right) = -\frac{1}{2}\left\{ p \log(2\pi) + \log\left|\mathbf{T}^2 + \mathbf{V}_i\right| + (\mathbf{r}_i - \rho_{\text{Random}})^{\mathrm{T}}\left(\mathbf{T}^2 + \mathbf{V}_i\right)^{-1}(\mathbf{r}_i - \rho_{\text{Random}})\right\}, \tag{6}$$

where p is the number of elements in $\mathbf{r}_i$ (Hardy & Thompson, 1996). The parameter estimates are obtained by maximizing the sum of the log-likelihoods of all studies. Since the dimensions of $\mathbf{r}_i$ may vary across studies, missing effect sizes are handled automatically. One advantage of using the MLE estimation method is that the parameter estimates are unbiased and efficient when the missingness is either MCAR or MAR.

When conducting a meta-analysis, we may quantify the degree of heterogeneity in effect sizes by the use of I^2, proposed by Higgins and Thompson (2002). The general formula is

$$I^2 = \frac{\widehat{\tau}^2}{\widehat{\tau}^2 + \widetilde{v}}, \tag{7}$$

where $\widehat{\tau}^2$ is the estimated heterogeneity and $\widetilde{v}$ is the "typical" within-study variance. The I^2 can be interpreted as the proportion of the total variation of the effect size that is due to the between-study heterogeneity. Higgins and Thompson proposed to estimate $\widetilde{v}$ as

$$\widetilde{v} = \frac{(n-1)\sum_{i=1}^{n} 1/v_i}{\left(\sum_{i=1}^{n} 1/v_i\right)^2 - \sum_{i=1}^{n} 1/v_i^2} \tag{8}$$

where n is the number of studies. As a rule of thumb, I^2s of 25 %, 50 %, and 75 % can be considered to indicate low, moderate, and high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). A multivariate version of I^2 has been developed by Jackson, White, and Riley (2012). An alternative approach is to calculate an I^2 on each of the effect sizes (correlation coefficients) in MASEM. That is, there will be five I^2s when five correlation coefficients are in the pooled correlation matrix. This gives an idea on how the heterogeneity spreads across the correlation elements.

Stage 2 analysis After the Stage 1 analysis with a random-effects model, a vector of the pooled correlation matrix $\widehat{\rho}_{\text{Random}}$ and its asymptotic sampling covariance matrix $\widehat{\mathbf{V}}_{\text{Random}}$ are estimated. The discrepancy function is the same as that under a fixed-effects model—that is,

$$F\left(\widehat{\gamma}\right) = \left(\rho_{\text{Random}} = -\rho\left(\widehat{\gamma}\right)\right)^{\mathrm{T}} \mathbf{V}_{\text{Random}}^{-1}\left(\rho_{\text{Random}} = -\rho\left(\widehat{\gamma}\right)\right) \tag{9}$$

As in Eq. 3, ρ_{Random} and $\mathbf{V}_{\text{Random}}$ are treated as observed values in Eq. 9. It should also be noted that the estimated variance component $\widehat{\mathbf{T}}^2$ does not directly enter into the discrepancy function in the Stage 2 analysis. Since $\mathbf{V}_{\text{Random}}$

is estimated after controlling for the random effects, it has already taken the random effects into account. Thus, $\mathbf{V}_{Random}$ is usually larger than $\mathbf{V}_{Fixed}$ from a fixed-effects model. The *SE*s of the parameter estimates from a random-effects model are also usually larger than those from a fixed-effects model.

Illustrations with the `metaSEM` package

The fixed- and random-effects TSSEM discussed in this article was implemented in the `metaSEM` package. This section demonstrates how to conduct the analyses in `R`. The `tssem1()` and `tssem2()` functions are used to conduct the Stage 1 and Stage 2 analyses, respectively. To conduct a fixed-effects TSSEM, users may specify the `method = "FEM"` argument in `tssem1()`, whereas a random-effects TSSEM may be specified via the `method = "REM"` argument. When categorical variables are present, the `cluster` argument may be used to specify a fixed-effects TSSEM on each group.

The `tssem2()` function is used to fit structural equation models in the Stage 2 analysis. This function automatically handles whether a fixed- or a random-effects model is used in the Stage 1 analysis. When a pooled correlation matrix is used to fit structural equation models, it is important to ensure that the diagonals of the model-implied correlation matrix are fixed at 1. This nonlinear constraint can be imposed by specifying the `diag.constraint = TRUE` argument in the `tssem2()` function. When nonlinear constraints are present, *SE*s are not reported in `OpenMx` since the *SE*s are not accurate. Likelihood-based CIs (LBCI) are generally recommended as a better alternative to CIs based on the *SE*s (Cheung, 2009). We may request the LBCI by specifying the `intervals = "LB"` argument in the `tssem2()` function.

The structural models in the Stage 2 analysis are specified via the reticular action model (RAM) formulation (McArdle & McDonald, 1984). The RAM model involves three matrices: **A**, **S**, and **F**. The **A** matrix specifies the asymmetric path (regression coefficients) from the independent variables to the dependent variables, whereas the **S** matrix specifies the symmetric paths (variances and covariances) of the variables. The **F** matrix is used to select the observed variables.

Two sample data sets, from Digman (1997) and B. J. Becker and Schram (1994), are used below to illustrate the proposed procedures on confirmatory factor analysis (CFA) and multiple regression analysis, respectively. These examples should be general enough for readers to be able to extend the techniques to more complicated models. The updated R code, outputs, and explanations are available at the author's website.

Data set from Digman (1997)

Method

Digman (1997) reported 14 correlation matrices among the five-factor model. He proposed that agreeableness (A), conscientiousness (C), and emotional stability (ES) were loaded under a higher-order factor, called "Alpha," that represents socialization, whereas extraversion (E) and intellect (I) were loaded under another higher-order factor, called "Beta," that represents personal growth. Figure 1 depicts the higher-order model with the corresponding elements labeled in the RAM formulation. The sample sizes of these 14 studies vary from 70 to 1,040. In Digman's article, he further grouped the studies under younger versus older participants. The data set was stored as `Digman97` in the `metaSEM` package, where the correlation matrices and sample sizes were stored as `Digman97$data` and `Digman97$n` respectively.

To conduct the Stage 1 analysis with a fixed-effects model, we may use the following syntax:

`fixed1 <- tssem1(my.df = Digman97$data, n = Digman97$n, method = "FEM"); summary(fixed1)`

After specifying the two-factor model via the RAM formulation—for example, **A1**, **S1**, and **F1**—we may use the following syntax to conduct the Stage 2 analysis:

`fixed2 <- tssem2(fixed1, Amatrix = A1, Smatrix = S1, Fmatrix = F1, diag.constraints = TRUE, intervals = "LB"); summary(fixed2)`

Results

In this section, the results for the fixed-effects TSSEM are presented by pooling all studies together. Since the homogeneity of the correlation matrices is questionable, I further present the results from treating the sample type as a categorical moderator (see Cheung & Chan, 2005a). Finally, I report the results for the random-effects TSSEM. Because
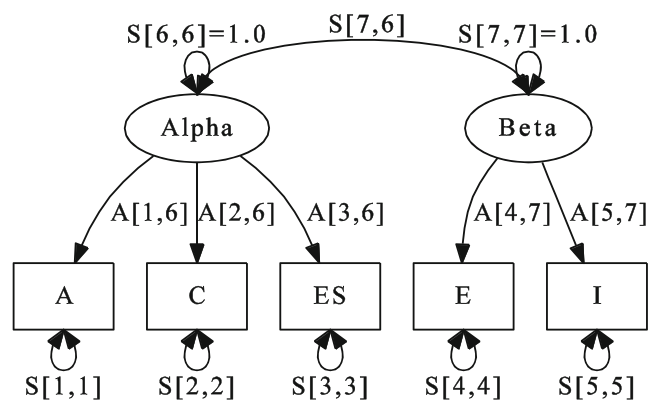

Fig. 1 A higher-order five-factor model

the results based on the fixed-effects models are also questionable, I will only interpret the parameter estimates on the basis of the random-effects model. Results related to the similarities and differences between the fixed- and random-effects models are discussed as well.

Fixed-effects model The goodness-of-fit indices for the Stage 1 analysis based on a fixed-effects TSSEM approach were $\chi^2(130, N = 4,496) = 1,499.73$, $p < .001$, CFI = 0.6825, RMSEA = 0.1812, and SRMR = 0.1750. On the basis of the test statistic and the goodness-of-fit indices, the assumption of homogeneity of the correlation matrices was rejected. As an illustration, I continued to fit the structural model on the basis of the pooled correlation matrix. The goodness-of-fit indices for the Stage 2 analysis were $\chi^2(4, \quad N = 4,496) = 65.06$, $p < .001$, CFI = 0.9802, RMSEA = 0.0583, and SRMR = 0.0284. The proposed model fits the data well.

Since the structural model was fitted on the basis of the pooled correlation matrix and its asymptotic covariance matrix, whether the correlation matrices were homogeneous had little impact on the model fit of the structural models. Readers should be cautious in interpreting the results of the Stage 2 analysis when the homogeneity of the correlation matrices is rejected in the Stage 1 analysis.

Fixed-effects model with a categorical moderator The 14 studies were grouped into the results from older and younger participants (`sample` in R). We may use the following code to conduct the Stage 1 and Stage 2 analyses on the older and younger participants:

```
fixed1.cluster <- tssem1(Digman97$data,
Digman97$n, method = "FEM", cluster = sample);
summary(fixed1.cluster); fixed2.cluster <-
tssem2(fixed1.cluster, Amatrix = A1, Smatrix =
S1, Fmatrix = F1, diag.constraints = TRUE,
intervals.type = "LB"); Summary(fixed2.cluster)
```

The goodness-of-fit indices of the Stage 1 analysis for the older and younger participants were $\chi^2(80, N = 3,658) = 823.88$, $p < .001$, CFI = 0.7437, RMSEA = 0.1513, and SRMR = 0.1528, and $\chi^2(40, N = 838) = 344.18$, $p < .001$, CFI = 0.7845, RMSEA = 0.2131, and SRMR = 0.1508, respectively. The assumption of the homogeneity of the correlation matrices in these two samples was rejected.

As an illustration, the structural models were still fitted. The goodness-of-fit indices of the Stage 2 analysis for the older and younger participants were $\chi^2(4, N = 3,658) = 21.92$, $p < .001$, CFI = 0.9921, RMSEA = 0.0350, and SRMR = 0.0160, and $\chi^2(4, N = 838) = 144.87$, $p < .001$, CFI = 0.9427, RMSEA = 0.2051, and SRMR = 0.1051, respectively. The proposed model appears to fit the data well in the older participants, but not in the younger participants. However, it should be noted again that the fit indices ignored the rejection of the homogeneity of the correlation matrices in

the Stage 1 analysis. Moreover, two improper solutions were obtained (an estimated factor loading of 3.28 and an estimated error variance of –9.82) in the younger participants. Thus, the results for the younger participants should not be trusted.

Random-effects model Since five variables were present, this led to a total of ten correlation coefficients in each study. If a random-effects model was fitted, the variance component of the random effects would include 55 elements. As not enough data were present to estimate the full variance component, a diagonal matrix was used to estimate the variance component. This option can be requested by specifying the `RE.type = "Diag"` argument in `tssem1()`. The Stage 1 and Stage 2 random-effects TSSEM can be conducted via the following commands:

```
random1 <- tssem1(Digman97$data, Digman97$n,
method="REM",RE.type="Diag");summary(random1);
random2 <- tssem2(random1, Amatrix = A1,
Smatrix = S1, Fmatrix = F1, diag.constraints =
TRUE, intervals = "LB")
```

It should be noted that goodness-of-fit indices are not available in the Stage 1 analysis under the random-effects model. The range of the I^2 index—the proportion of total variance that could be explained by the between-study effect—was from .84 to .95. This indicates huge between-study heterogeneity. A random-effects model is therefore preferred to a fixed-effects model.

The pooled correlation matrix based on the random-effects model was used to fit the two-factor CFA in the Stage 2 analysis. The goodness-of-fit indices for the proposed model were $\chi^2(4, N=4,496) = 8.51$, $p < .001$, CFI = 0.9776, RMSEA = 0.0158, and SRMR = 0.0463. The proposed model fits the data very well.

The results support Digman's (1997) higher-order factor structure. All standardized factor loadings were very high, ranging from 0.57 to 0.77. There is one difference between Digman's proposal and the present findings, however: Digman argued that the factors were uncorrelated because they were the fundamental factors. The results show that the two factors were moderately correlated, with a correlation of .39 and a 95 % LBCI of .30 to .49.

Data set from B. J. Becker and Schram (1994)

Method

B. J. Becker and Schram (1994) used five samples to measure the correlations among SAT (math), SAT (verbal), and spatial ability. Since B. J. Becker and Schram divided the data between male and female participants, this led to a total of ten independent samples. B. J. Becker and Schram synthesized the correlation matrices and fitted a regression model by using SAT (math) as the dependent variable and SAT (verbal) and spatial

ability as the predictors. Figure 2 shows the regression model with the corresponding elements in the RAM formulation. The sample sizes vary from 18 to 153. The data set was stored as `Becker94` in the `metaSEM` package. The syntax to conduct the analyses was similar to the examples listed for Digman's (1997) data set; they are not repeated here. Readers may refer to the online supplementary materials for the details http://courses.nus.edu.sg/course/psycwlm/Internet/metaSEM/masem.html.

Results

First, the results for the fixed-effects TSSEM will be presented by pooling all studies together and treating gender as a categorical moderator. Then I present the results for the random-effects TSSEM. Since the random-effects model was more appropriate, I only interpret the parameter estimates based on the random-effects model.

Fixed-effects model The goodness-of-fit indices of the Stage 1 analysis were $\chi^2(27, N = 538) = 62.50$, $p < .001$, CFI = 0.7943, RMSEA = 0.1565, and SRMR = 0.2011. On the basis of the test statistic and the goodness-of-fit indices, the hypothesis of the homogeneity of the correlation matrices was rejected. As an illustration, I will continue to fit the structural model on the basis of the pooled correlation matrix. Since the regression model is saturated, the test statistic is 0.

Fixed-effects model with a categorical moderator The ten studies were grouped into studies with female and male participants. The goodness-of-fit indices of the Stage 1 analysis for the female and male participants were $\chi^2(12, N = 235) = 42.41$, $p < .001$, CFI = 0.7116, RMSEA = 0.2327, and SRMR = 0.2339, and $\chi^2(12, N = 303) = 16.13$, $p = .1852$, CFI = 0.9385, RMSEA = 0.0755, and SRMR = 0.1508,
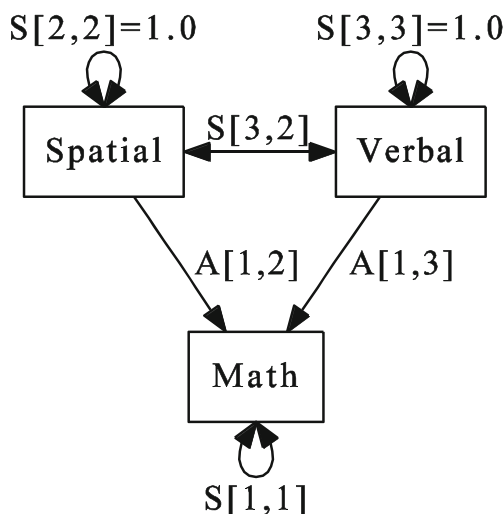
Fig. 2 A multiple regression on mathematical performance

respectively. The homogeneity of correlation matrices seem adequate in the male participants, but not in the female participants. Since the sample sizes were not large, we should be cautious in concluding that the data were homogeneous. As an illustration, I continued to fit the structural models.

Random-effects model Since three variables were present, this led to a total of three correlation coefficients in each study. If a random-effects model were fitted, a total of 6 elements would be included in the variance component of the random effects. Since there were only ten studies, a diagonal matrix was used to fit the variance component.

The I^2 indices for the correlations between spatial and math, verbal and math, and spatial and verbal were .00, .81, and .23, respectively. These indicate that the heterogeneity in the correlation between spatial and math was trivial, whereas the heterogeneity was largest in the correlation between verbal and math. A random-effects model was therefore preferred to a fixed-effects model. The regression coefficients (and their 95 % LBCIs) from spatial and verbal to math were 0.30 (0.21, 0.37) and 0.37 (0.21, 0.53), respectively. The error variance for math was .73, meaning that the R^2 was (1 − .73) or .27.

Discussion

The parameter estimates, their 95 % LBCIs, and the widths of the CIs are listed in Tables 1 and 2. As we can see from the tables, there is no systematic relationship between the parameter estimates from the fixed- and the random-effects models. However, the CIs of the random-effects model are usually larger than those of the fixed-effects model. The *SE*s and CIs of the fixed-effects model are underestimated when the assumption of the homogeneity of effect sizes is not valid. The same findings have always been observed in conventional meta-analysis (e.g., Hedges & Vevea, 1998).

Another observation is that the Stage 2 analysis fits the proposed models well, even when the assumption of homogeneity of the correlation matrices is clearly violated in the Stage 1 analysis. This occurs because the Stage 2 analysis mainly uses the pooled correlation matrix as the input. Information on the heterogeneity of the correlation matrices is not incorporated under the fixed-effects model. Researchers should be cautioned that the fit indices in the Stage 2 analysis may be misleading if a fixed-effects model is incorrectly applied to a pool of heterogeneous correlation matrices. This article has demonstrated how a random-effects TSSEM can be applied when the assumption of homogeneity of the correlation matrices is not met.

Conclusion and future directions

This article has reviewed the basic ideas of MASEM. More specifically, the TSSEM approach was introduced and

Table 1 Parameter estimates and their 95 % likelihood-based confidence intervals in the Stage 2 analyses of Digman (1997)

	Estimate	Lower Bound	Upper Bound	Width of CI
	Fixed-Effects Model			
Amatrix[1,6] A loaded on Alpha	0.5626	0.5324	0.5929	0.0604
Amatrix[2,6] C loaded on Alpha	0.6051	0.5751	0.6353	0.0602
Amatrix[3,6] ES loaded on Alpha	0.7191	0.6886	0.7503	0.0617
Amatrix[4,7] E loaded on Beta	0.7820	0.7191	0.8559	0.1368
Amatrix[5,7] I loaded on Beta	0.5509	0.4994	0.6023	0.1029
Smatrix[1,1] Error variance of A	0.6835	0.6485	0.7165	0.0680
Smatrix[2,2] Error variance of C	0.6338	0.5964	0.6693	0.0729
Smatrix[3,3] Error variance of ES	0.4829	0.4370	0.5258	0.0888
Smatrix[4,4] Error variance of E	0.3885	0.2674	0.4829	0.2155
Smatrix[5,5] Error variance of I	0.6965	0.6372	0.7506	0.1134
Smatrix[7,6] Factor correlation between Alpha and Beta	0.3626	0.3184	0.4065	0.0881
	Fixed-Effects Model for Older Participants			
Amatrix[1,6] A loaded on Alpha	0.5125	0.4769	0.5484	0.0715
Amatrix[2,6] C loaded on Alpha	0.5500	0.5149	0.5854	0.0705
Amatrix[3,6] ES loaded on Alpha	0.7321	0.6956	0.7701	0.0745
Amatrix[4,7] E loaded on Beta	0.9675	0.8680	1.1096	0.2416
Amatrix[5,7] I loaded on Beta	0.4305	0.3692	0.4869	0.1177
Smatrix[1,1] Error variance of A	0.7373	0.6993	0.7726	0.0733
Smatrix[2,2] Error variance of C	0.6974	0.6573	0.7349	0.0776
Smatrix[3,3] Error variance of ES	0.4640	0.4070	0.5162	0.1092
Smatrix[4,4] Error variance of E	0.0639	−0.2320	0.2466	0.4785
Smatrix[5,5] Error variance of I	0.8147	0.7629	0.8637	0.1008
Smatrix[7,6] Factor correlation between Alpha and Beta	0.3491	0.2921	0.4032	0.1112
	Fixed-Effects Model for Younger Participants			
Amatrix[1,6] A loaded on Alpha	0.7476	0.7007	0.7947	0.0940
Amatrix[2,6] C loaded on Alpha	0.9120	0.8732	0.9515	0.0783
Amatrix[3,6] ES loaded on Alpha	0.6772	0.6260	0.7275	0.1015
Amatrix[4,7] E loaded on Beta	0.1524	0.0154	0.3417	0.3263
Amatrix[5,7] I loaded on Beta	3.2898	1.5252	209.3000	207.7748
Smatrix[1,1] Error variance of A	0.4410	0.3684	0.5091	0.1406
Smatrix[2,2] Error variance of C	0.1683	0.0946	0.2376	0.1430
Smatrix[3,3] Error variance of ES	0.5414	0.4707	0.6081	0.1374
Smatrix[4,4] Error variance of E	0.9768	0.8833	0.9994	0.1161
Smatrix[5,5] Error variance of I	−9.8226	−488,900	−1.3113	488,899
Smatrix[7,6] Factor correlation between Alpha and Beta	0.1171	0.0108	0.2750	0.2642
	Random-Effects Model			
Amatrix[1,6] A loaded on Alpha	0.5726	0.4737	0.6769	0.2032
Amatrix[2,6] C loaded on Alpha	0.5901	0.4905	0.6949	0.2044
Amatrix[3,6] ES loaded on Alpha	0.7705	0.6599	0.9043	0.2444
Amatrix[4,7] E loaded on Beta	0.6934	0.5626	0.8718	0.3093
Amatrix[5,7] I loaded on Beta	0.6401	0.5083	0.7864	0.2781
Smatrix[1,1] Error variance of A	0.6722	0.5418	0.7756	0.2339
Smatrix[2,2] Error variance of C	0.6518	0.5171	0.7594	0.2423
Smatrix[3,3] Error variance of ES	0.4064	0.1819	0.5645	0.3826
Smatrix[4,4] Error variance of E	0.5192	0.2394	0.6835	0.4442
Smatrix[5,5] Error variance of I	0.5903	0.3813	0.7416	0.3603
Smatrix[7,6] Factor correlation between Alpha and Beta	0.3937	0.3024	0.4903	0.1879

The Amatrix and Smatrix refer to the asymmetric matrix of regression coefficients and the symmetric matrix of variance and covariance elements, respectively. Please refer to Fig. 1 for the corresponding elements.

Table 2 Parameter estimates and their 95 % likelihood-based confidence intervals in the Stage 2 analyses of B. J. Becker and Schram (1994)

	Estimate	Lower Bound	Upper Bound	Width of CI
	Fixed-Effects Model			
Amatrix[1,2] Regression coefficient from spatial to math	0.3315	0.2577	0.4052	0.1475
Amatrix[1,3] Regression coefficient from verbal to math	0.2740	0.1966	0.3515	0.1550
Smatrix[1,1] Error variance on math	0.7831	0.7155	0.8424	0.1269
Smatrix[2,3] Correlation between spatial and verbal	0.1756	0.0918	0.2595	0.1677
	Fixed-Effects Model for Female Participants			
Amatrix[1,2] Regression coefficient from spatial to math	0.4103	0.3051	0.5147	0.2096
Amatrix[1,3] Regression coefficient from verbal to math	0.2691	0.1525	0.3859	0.2334
Smatrix[1,1] Error variance on math	0.7217	0.6131	0.8144	0.2013
Smatrix[2,3] Correlation between spatial and verbal	0.1705	0.0420	0.2989	0.2569
	Fixed-Effects Model for Male Participants			
Amatrix[1,2] Regression coefficient from spatial to math	0.2677	0.1659	0.3695	0.2037
Amatrix[1,3] Regression coefficient from verbal to math	0.2783	0.1750	0.3815	0.2065
Smatrix[1,1] Error variance on math	0.8242	0.7367	0.8953	0.1586
Smatrix[2,3] Correlation between spatial and verbal	0.1792	0.0685	0.2899	0.2213
	Random-Effects Model			
Amatrix[1,2] Regression coefficient from spatial to math	0.2954	0.2144	0.3741	0.1597
Amatrix[1,3] Regression coefficient from verbal to math	0.3717	0.2131	0.5296	0.3165
Smatrix[1,1] Error variance on math	0.7301	0.5903	0.8305	0.2402
Smatrix[2,3] Correlation between spatial and verbal	0.2029	0.1118	0.2942	0.1824

The Amatrix and Smatrix refer to the asymmetric matrix of regression coefficients and the symmetric matrix of variance and covariance elements, respectively. Please refer to Fig. 2 for the corresponding elements.

extended to a random-effects model. The illustrations demonstrated how to conduct the fixed- and random-effects TSSEM using the `metaSEM` package. This article has only demonstrated the functions in the `metaSEM` package that are related to TSSEM. In the package are other useful functions for general meta-analysis using an SEM approach—for example, univariate and multivariate meta-analysis (Cheung, 2008, in press-b) and three-level meta-analysis (Cheung, in press-a). Since a SEM approach is used to model meta-analytic data, flexible constraints on the parameter estimates may be imposed. Besides the conventional ML estimation method, restricted maximum likelihood (REML) may also be used to conduct the analyses (Cheung, 2013a).

Since the common objective of a MASEM is to generalize findings beyond the studies included in a meta-analysis, the random-effects TSSEM proposed in this article should generally be recommended. Applications and methodological development using a random-effects model are still limited in the MASEM literature, and several open questions deserve attention for future research. I will address some of them below.

Distribution assumptions for the data

For most data analyses in the social and behavioral sciences, the normality of the data is usually assumed for ease of analysis. When there are reasons to question the validity of the

normality assumption, robust statistics may be used to adjust for the nonnormality effect. Robust statistics are available in multilevel modeling (Litière, Alonso, & Molenberghs, 2008; Verbeke & Lesaffre, 1997), SEM (Yuan & Bentler, 2007), and meta-analysis (Sidik & Jonkman, 2006). Applying robust statistics to MASEM, however, is more complicated. As is shown in Eq. 5, MASEM has two variance components—V_i, the known sampling covariance matrix in the ith study, and T^2, the variance component of the study-specific random effects. Since raw data are usually not available in MASEM, the use of the WLS estimation method in TSSEM does not automatically correct for nonnormality. The effects of nonnormality on V_i and T^2 and how to apply the robust statistics in MASEM still remain unknown. Further research should address the effects of nonnormality in MASEM.

Model selections and comparisons with fit indices

The proposed model follows a chi-square distribution only if (1) the proposed model is correct; (2) the sample size is large enough; and (3) the distribution assumption is correct. When the proposed model is rejected, this could be attributed either to model misspecification or to other violations of assumptions (e.g., Saris, Satorra, & van der Veld, 2009). Instead of relying on the formal test statistic, many goodness-of-fit indices—such as CFI, TLI, and RMSEA—have been

proposed to assess the approximate fit of the proposed model (see Barrett, 2007, and the commentaries regarding the arguments for and against the goodness-of-fit indices). Since TSSEM is a special case of SEM, the arguments for and against the goodness-of-fit indices can be directly applied to TSSEM. Several issues require special attention, however: First, TSSEM is usually driven by comparing several theoretically meaningful models. Therefore, instead of evaluating whether a particular model is appropriate for the data, a better approach would be only to compare the relative fits of a number of different a priori models. Second, TSSEM is usually based on a large set of divergent samples and a much larger overall sample size than is conventional SEM. Thus, it is likely that the large-sample requirement will not be difficult to satisfy in the majority of TSSEM applications. It will be critical for future work to study how best to evaluate model specification in the context of TSSEM.

Bayesian approach to MASEM

As discussed above, one of the limitations in model testing in SEM is that the proposed model may not be exactly correct. Using Digman's (1997) higher-order factor structure as an example, the proposed model hypothesizes no double loading on the items and that the measurement errors are uncorrelated. Under the conventional SEM framework, it is not possible to free all of the possible loadings and covariances among the error variances. Muthén and Asparouhov (2012; see also the commentaries in that special issue) argued that a Bayesian approach might be used to address some of these concerns through the use of informative priors. As Bayesian statistics are becoming more and more popular in SEM (Lee, 2007; Muthén & Asparouhov, 2012) and meta-analysis (Smith, Spiegelhalter, & Thomas, 1995; Sutton & Abrams, 2001), it may be feasible to apply a Bayesian approach to MASEM. For example, Prevost et al. (2007) showed how the Bayesian approach could be used to synthesize correlation matrices with a random-effects model. Since MASEM consists of two stages of analyses, further research would investigate how the Bayesian approach could be integrated into the MASEM framework.

Individual patient data meta-analysis

In the area of medical sciences, the individual patient data meta-analysis—that is, the synthesis of raw data rather than summary statistics—is becoming popular (e.g., Riley, Lambert, & Abo-Zaid, 2010; Stewart et al., 2012; Sutton, Kendrick, & Coupland, 2008). A similar approach, termed *integrative data analysis*, has also been proposed in psychology (see Curran, 2009, in the special issue of *Psychological Methods*). Many of the aforementioned issues in MASEM can be addressed when raw data are available. Moreover, the response variables can be continuous, categorical, count,

nominal, or a mix of these options. This will broaden the usefulness of MASEM. The main obstacle, however, is that raw data are usually not made publicly available in psychology and in the behavioral sciences in general.

Inclusion of study characteristics

Study characteristics are usually included as predictors in meta-analysis. The study characteristics may be used to explain the heterogeneity of effect sizes. This is known as *mixed-effects meta-analysis* or *meta-regression* (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009). In the illustrations above, the studies were classified into two groups for separate analyses. This approach, however, is limited to categorical variables only. When the variables are continuous—for example, year of publication and duration of intervention—reviewers may have to categorize them into groups. This practice is not optimal, because information will be lost in the categorization process.

Since two stages of analyses are used in TSSEM, continuous covariates may be used in the Stage 1 or Stage 2 analyses. A multivariate meta-analysis may be conducted in the Stage 1 analysis by using study characteristics as the predictors. However, the pooled correlation or covariance matrix depends on the values of the study characteristics. It is not clear how the Stage 2 model could be fitted after the Stage 1 analysis. If the Stage 1 analysis is conducted without the study characteristics, it is still unclear how the study characteristics could be used to model the Stage 2 analysis. Further research may address how continuous study characteristics can be used in MASEM.

Correcting for unreliability

There is some controversy on whether it is necessary to correct for attenuations or statistical artifacts in meta-analysis. Since the measurements are liable to measurement errors, the observed correlation coefficients are usually smaller than the actual correlations. Rosenthal (1991) criticizes the use of correction for attenuations because the corrected values are different from the "typical research findings" and the corrected values are not as useful as uncorrected values in realistic settings. Other researchers—for example, Hunter and Schmidt (2004), prefer to correct for attenuation before combining them, whereas others suggest that combining the observed correlation coefficients is sufficient.

Hunter and Schmidt (2004) identified 11 artifacts that could be corrected before combining the correlation coefficients. These included sampling error, error of measurement in the dependent and independent variables, restriction of range, and so forth. However, it is unlikely that the published articles would include all of the information necessary for correction.

One type of measurement error is unreliability, which can be corrected by the equation

$$r_{Corrected} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}}, \tag{10}$$

where $r_{Corrected}$ is the estimated corrected correlation for the unreliability of measurements, r_{xy} is the observed correlation between variables x and y, and r_{xx} and r_{yy} are the estimated reliabilities of variables x and y, respectively.

When item-level data are available, there is no need to correct for the unreliability in MASEM. This is because both CFA and SEM on the item-level data can account for the measurement errors. When MASEM is conducted on the composite scores, reviewers may need to decide whether to apply the correction. By reviewing several published meta-analyses, Michel, Viswesvaran, and Thomas (2011) recently argued that substantive model conclusions are generally unaffected by study artifacts and the related statistical corrections in the psychological literature. Since Michel et al.'s conclusions were based on real examples rather than on computer simulation, further research may need to address the effects of unreliability in MASEM.

To conclude, the TSSEM provides a valuable approach for researchers conducting MASEM. Fixed- and the random-effects models are based on different assumptions. The implementation of these techniques in the package makes both techniques accessible to applied researchers, who may then choose the correct model to fit their research settings and research questions.

References

Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological Methodology, 17,* 71–103.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815–824.

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics, 17,* 341–362.

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

Becker, G. (1996). The meta-analysis of factor analyses: An illustration based on the cumulation of correlation matrices. *Psychological Methods, 1,* 341–353.

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: Sage.

Becker, B. J., & Schram, C. M. (1994). Examining explanatory models through research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 357–381). New York, NY: Sage.

Beretvas, S., & Furlow, C. (2006). Evaluation of an approximate method for synthesizing covariance matrices for use in meta-analytic SEM. *Structural Equation Modeling, 13,* 153–185.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., & Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76,* 306–317.

Bonett, D. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods, 14,* 225–238.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Hoboken, NJ: Wiley.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

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.

Cheung, M. W.-L. (2009). Constructing approximate confidence intervals for parameters with structural equation models. *Structural Equation Modeling, 16,* 267–294.

Cheung, M. W.-L. (2010). Fixed-effects meta-analyses as multiple-group structural equation models. *Structural Equation Modeling, 17,* 481–509.

Cheung, M. W.-L. (2013a). Implementing restricted maximum likelihood estimation in structural equation models. *Structural Equation Modeling, 20,* 157–167.

Cheung, M. W.-L. (2013b). *metaSEM: Meta-analysis using structural equation modeling*. Retrieved from http://courses.nus.edu.sg/course/psycwlm/Internet/metaSEM/

Cheung, M. W.-L. (in press-a). Modeling dependent effect sizes with three-level meta-analysis: A structural equation modeling approach. *Psychological Methods*.

Cheung, M. W.-L. (in press-b). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling*.

Cheung, M. W.-L., & Chan, W. (2004). Testing dependent correlation coefficients via structural equation modeling. *Organizational Research Methods, 7,* 206–223.

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

Cheung, M. W.-L., & Chan, W. (2005b). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10,* 40–64.

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

Cheung, M. W.-L., Leung, K., & Au, K. (2006). Evaluating multilevel models in cross-cultural research: An illustration with Social Axioms. *Journal of Cross-Cultural Psychology, 37,* 522–541.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85,* 678–707.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84,* 3–13.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Sage.

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

Curran, P. J. (2009). The seemingly quixotic pursuit of a cumulative psychological science: Introduction to the special issue. *Psychological Methods, 14,* 77–80.

Curran, P., West, S., & Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

Development Core Team, R. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

Eby, L. T., Freeman, D. M., Rush, M. C., & Lance, C. E. (1999). Motivational bases of affective organizational commitment: A partial test of an integrative theoretical model. *Journal of Occupational and Organizational Psychology, 72,* 463–483.

Hardy, R. J., & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine, 15,* 619–629.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L., & Vevea, J. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560.

Hom, P. W., Caranikas-Walker, F., Prussia, G. E., & Griffeth, R. W. (1992). A meta-analytical structural equations analysis of a model of employee turnover. *Journal of Applied Psychology, 77,* 890–909.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Jackson, D., White, I. R., & Riley, R. D. (2012). Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Statistics in Medicine, 31,* 3805–3820.

Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: Recommendations and strategies. *Journal of Business and Psychology.* doi:10.1007/s10869-013-9285-x. Advance online publication.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester, UK: Wiley.

Litière, S., Alonso, A., & Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine, 27,* 3125–3144.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 234–251.

Michel, J. S., Viswesvaran, C., & Thomas, J. (2011). Conclusions from meta-analytic structural equation models generally do not change due to corrections for study artifacts. *Research Synthesis Methods, 2,* 174–187.

Murayama, K., & Elliot, A. J. (2012). The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin, 138,* 1035–1070.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17,* 313–335.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 51,* 431–462.

Prevost, A., Mason, D., Griffin, S., Kinmonth, A., Sutton, S., & Spiegelhalter, D. (2007). Allowing for correlations between correlations in random-effects meta-analysis of correlation matrices. *Psychological Methods, 12,* 434–450.

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal, 340,* c221.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Revth ed.). Newbury Park, CA: Sage.

Saris, W. E., Satorra, A., & van der Veld, W. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16,* 561–582.

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

Shuster, J. J. (2010). Empirical vs. natural weighting in random effects meta-analysis. *Statistics in Medicine, 29,* 1259–1265.

Sidik, K., & Jonkman, J. N. (2006). Robust variance estimation for random effects meta-analysis. *Computational Statistics and Data Analysis, 50,* 3681–3701.

Smith, T. C., Spiegelhalter, D. J., & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine, 14,* 2685–2699.

Steinmetz, H., Baeuerle, N., & Isidor, R. (2012). Testing the circular structure of human values: A meta-analytical structural equation modeling approach. *Survey Research Methods, 6,* 61–75.

Stewart, G. B., Altman, D. G., Askie, L. M., Duley, L., Simmonds, M. C., & Stewart, L. A. (2012). Statistical analysis of individual participant data meta-analyses: A comparison of methods and recommendations for practice. *PLoS ONE, 7,* e46042.

Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research, 10,* 277–303.

Sutton, A. J., Kendrick, D., & Coupland, C. A. C. (2008). Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine, 27,* 651–669.

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology, 46,* 259–290.

Verbeke, G., & Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis, 23,* 541–556.

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

Yuan, K.-H., & Bentler, P. M. (2007). Robust procedures in structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 367–397). Amsterdam, The Netherlands: Elsevier/North-Holland.