Psychological Methods
2008, Vol. 13, No. 3, 182–202

1082-989X/08/$12.00 DOI: 10.1037/a0013163

A Model for Integrating Fixed-, Random-, and Mixed-Effects Meta-Analyses Into Structural Equation Modeling

Mike W.-L. Cheung
National University of Singapore

Meta-analysis and structural equation modeling (SEM) are two important statistical methods in the behavioral, social, and medical sciences. They are generally treated as two unrelated topics in the literature. The present article proposes a model to integrate fixed-, random-, and mixed-effects meta-analyses into the SEM framework. By applying an appropriate transformation on the data, studies in a meta-analysis can be analyzed as subjects in a structural equation model. This article also highlights some practical benefits of using the SEM approach to conduct a meta-analysis. Specifically, the SEM-based meta-analysis can be used to handle missing covariates, to quantify the heterogeneity of effect sizes, and to address the heterogeneity of effect sizes with mixture models. Examples are used to illustrate the equivalence between the conventional meta-analysis and the SEM-based meta-analysis. Future directions on and issues related to the SEM-based meta-analysis are discussed.

Keywords: meta-analysis, structural equation model, fixed-effects model, random-effects model, mixed-effects model

It is of methodological importance to see how seemingly unrelated statistical methods can be linked together. Consider the classic example of analysis of variance (ANOVA) and multiple regression. Before the seminal work of Cohen (1968; Cohen & Cohen, 1975), "the textbooks in 'psychological' statistics treat[ed multiple regression, ANOVA, and ANCOVA] quite separately, with wholly different algorithms, nomenclature, output, and examples" (Cohen, 1968, p. 426). Understanding the mathematical equivalence between an ANOVA (and analysis of covariance, or ANCOVA) and a multiple regression helps us to comprehend the details behind the general linear model.

Another important example in social statistics has been the development of structural equation modeling (SEM; e.g., Bentler, 2004; Bollen, 1989; Jöreskog & Sörbom, 1996; L. K. Muthén & Muthén, 2007). SEM provides a flexible framework for testing complicated models involving latent and observed variables. It integrates ideas of latent variables in psychometrics, path models in sociology, and structural models in econometrics. The general linear model, path analysis, and confirmatory factor analysis are some special cases of SEM.

Recently, it has been shown that many models used in the social and behavioral sciences are related to SEM. Takane and de Leeuw (1987; see also MacIntosh & Hashim, 2003) showed how some item response theory (IRT) models could be analyzed as structural equation models. Several authors (e.g., Bauer, 2003; Curran, 2002; Mehta & Neale, 2005; Rovine & Molenaar, 2000, 2001) demonstrated how multilevel models could be formulated as structural equation models. The advantage of integrating these models together is that a unified framework can be used to address complex research questions involving some of these models. There are at least two such general models: Mplus (L. K. Muthén & Muthén, 2007) combines SEM, multilevel models, mixture modeling, survival analysis, latent class models, and some IRT models into a single statistical modeling framework, whereas Generalized Linear Latent and Mixed Models (GLLAMM; Skrondal & Rabe-Hesketh, 2004) integrates SEM, generalized linear models, multilevel models, latent class models, and IRT models.

Meta-analysis, a term coined by Glass (1976), represents "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating

Portions of this article were presented at the 29th International Congress of Psychology, Berlin, Germany, July 2008. This research was supported by the Academic Research Fund Tier 1 (R-581-000-064-112) from the Ministry of Education, Singapore. I thank Wolfgang Viechtbauer for suggesting references on the relationship between the maximum likelihood and restricted maximum likelihood estimates of the variance component and Maggie Chan for providing comments on a previous version of this article.

Correspondence concerning this article should be addressed to Mike 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. E-mail: mikewlcheung@nus.edu.sg

the findings" (p. 3). It is also known as research synthesis (Cooper & Hedges, 1994), the combining of information (National Research Council, 1992), and systematic review (Petticrew & Roberts, 2006). It has become the standard methodology in synthesizing research findings in the social, behavioral, and medical sciences.

Meta-analysis and SEM have their own traditions and terminologies. They are usually treated as two unrelated topics in the literature. For example, Stapleton and Leite (2005) reviewed 55 SEM course syllabi; none of them mentioned meta-analysis as a topic. Few SEM books discuss meta-analysis as a topic relevant to SEM.[1] In the literature on meta-analysis, Hunter and Schmidt (2004) briefly mentioned a procedure of combining meta-analytic and SEM techniques, which was called *meta-analytic structural equation modeling* (MASEM; M. W. L. Cheung & Chan, 2005, in press).[2] However, MASEM is limited to combining correlation matrices and covariance matrices. The pooled correlation or covariance matrix is then used to estimate a structural model. It is not intended to synthesize other effect sizes such as Hedges's (1981) *d*, the standardized mean change for repeated measures (Becker, 1988; Morris & DeShon, 2002), or the odds ratio (Haddock, Rindskopf, & Shadish, 1998).

The primary objective of this article is to propose a model to integrate meta-analysis into the SEM framework. Fixed-, random-, and mixed-effects meta-analyses (e.g., Hedges, 1994; Hedges & Olkin, 1985; Hedges & Vevea, 1998) can be formulated as structural equation models. In other words, SEM can be directly used to conduct a meta-analysis. For the sake of discussion, in this article this new approach is called *SEM-based meta-analysis*. This article contributes to the methodological development of a new area of research by integrating meta-analysis and SEM.

Apart from being of methodological interest, SEM-based meta-analysis also has several practical benefits for applied researchers who are conducting a meta-analysis. As many state-of-the-art techniques have been implemented in many SEM packages, these techniques are readily accessible to researchers by using the SEM-based meta-analysis. Specifically, this article will show how SEM-based meta-analysis can be used to handle missing covariates, to quantify the heterogeneity of effect sizes, and to address the heterogeneity of effect sizes with mixture models in a meta-analysis. Another benefit of the proposed approach is that mathematical models of meta-analysis can be easily translated into path diagrams in the SEM-based meta-analysis. Path diagrams preserve all of the necessary components in the mathematical models (Curran & Bauer, 2007). They provide an alternative and accurate presentation of the mathematical aspect of a meta-analysis. SEM-based meta-analysis may make meta-analytic techniques more accessible to applied researchers.

This article is organized as follows. The next section contains a brief review of the various meta-analytic models. A model for integrating meta-analysis into SEM is then presented. A data set from Hox (2002) is used to demonstrate the equivalence between conventional meta-analytic techniques and SEM-based meta-analysis. Finally, future directions on this approach and related issues are discussed.

Meta-Analytic Models

Fixed-Effects Models

Effect sizes are quantitative indices that summarize the results of a study. Common effect sizes include Hedges's *d*, the correlation coefficient (and its Fisher's *z* transformed score), and the odds ratio. Fleiss (1994) and Rosenthal (1994) have provided comprehensive summaries of many effect sizes and their estimated sampling variances.

Models without any covariate. The simplest analysis in a meta-analysis involves pooling a series of independent effect sizes using a fixed-effects model. In this article, I denote y_i as a generic effect size in the *i*th study. y_i is usually expressed as

$$y_i = \beta_{\text{Fixed}} + e_i, \tag{1}$$

where β_{Fixed} and e_i are the population effect size and the sampling error in the *i*th study, respectively. e_i is assumed to be normally distributed with a mean of zero and a known variance of σ_i^2.

[1] Skrondal and Rabe-Hesketh (2004) presented a unified latent variable modeling approach to multilevel, longitudinal, IRT, and structural equation models as well as many extensions. An example of meta-analysis was discussed in pp. 299–307. Conventional meta-analytic techniques (not based on SEM) were presented as the framework of analyses.

[2] MASEM can be illustrated by using an example from Becker (1992, 1995) of combining independent studies that addressed the predictability of mathematics aptitude from spatial ability and verbal ability. Correlation matrices among these variables from four studies were synthesized with the generalized least squares method. A linear model was then fitted on the pooled correlation matrix by using mathematical aptitude as the dependent variable and spatial ability and verbal ability as the independent variables. Besides fitting regression models on the pooled correlation matrix, MASEM has been extended to fit path models, exploratory factor analysis, confirmatory factor analysis, and structural equation models (e.g., Becker, 2000; Becker & Schram, 1994; S. F. Cheung, 2000; Furlow & Beretvas, 2005; Hafdahl, 2001; Shadish, 1996; Viswesvaran & Ones, 1995).

Conventionally, researchers have synthesized correlation matrices with meta-analytic techniques (e.g., Becker, 1992; Hedges & Olkin, 1985; Hunter & Schmidt, 2004), whereas the pooled correlation matrix is fitted in SEM. M. W. L. Cheung and Chan (2005, in press) proposed an SEM approach to conducting MASEM. Both synthesizing correlation matrices and fitting SEM on the pooled correlation matrix are conducted in the SEM framework.

The weighted effect size $\hat{\beta}_{\text{Fixed}}$ under the fixed-effects model is

$$\hat{\beta}_{\text{Fixed}} = \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i} \tag{2}$$

where $w_i = 1/\sigma_i^2$ is the weight and k is the total number of studies. The sampling variance $\hat{s}^2_{\text{Fixed}}$ of $\hat{\beta}_{\text{Fixed}}$ is computed by

$$\hat{s}^2_{\text{Fixed}} = 1/\sum_{i=1}^{k} w_i. \tag{3}$$

If the population effect sizes are homogeneous, $\hat{\beta}_{\text{Fixed}}$ is an unbiased estimate of the population effect size. Moreover, $\hat{\beta}_{\text{Fixed}}$ has the smallest sampling variance of all possible weighted estimators when the sampling variances are truly known. Because σ_i^2 is usually estimated, the accuracy of $\hat{s}^2_{\text{Fixed}}$ in estimating the sampling variance of $\hat{\beta}_{\text{Fixed}}$ depends on how accurate the estimated value of σ_i^2 is (see Hedges, 2007; Hedges & Olkin, 1985). After the averaging, it is of interest to test whether the weighted effect size $\hat{\beta}_{\text{Fixed}}$ is statistically significant. We may compute a test statistic

$$Z_1 = \hat{\beta}_{\text{Fixed}}/\hat{s}_{\text{Fixed}}. \tag{4}$$

Under the null hypothesis $H_0 : \beta_{\text{Fixed}} = 0$, the test statistic Z_1 has an approximate standard normal distribution.

Combining estimates of effect sizes across studies with the fixed-effects model is appropriate only when the effect sizes are homogeneous (National Research Council, 1992); otherwise, the estimated standard error on the weighted mean under the fixed-effects models is smaller than its true value when the effect sizes are heterogeneous. To test the homogeneity of the effect sizes, we may compute a Q statistic (Cochran, 1954):

$$Q = \sum_{i=1}^{k} w_i (y_i - \hat{\beta}_{\text{Fixed}})^2. \tag{5}$$

Under the null hypothesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k$, the Q statistic has an approximate chi-square distribution with $(k - 1)$ degrees of freedom. However, this does not necessary mean that fixed-effects models should never be used whenever the effect sizes are heterogeneous. Hedges and Vevea (1998) pointed out that fixed-effects models are still appropriate even if the effect sizes are heterogeneous when the researchers are interested in only this collection of studies. This is what they call a conditional inference.

Models with covariates. Besides estimating a common effect size, study characteristics may be used as covariates to model the variability among the effect sizes. Study characteristics can be in the form of categorical covariates (Hedges, 1982a) and continuous covariates (Hedges, 1982b). A weighted least squares (WLS) approach is usually used to model the variability among the effect sizes with covariates (e.g., Hedges & Olkin, 1985). It would be more convenient to express the model in matrix notation

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e}, \tag{6}$$

where $\mathbf{y}$ is a $k \times 1$ vector of effect sizes, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients including the intercept, $\mathbf{e}$ is a $k \times 1$ vector of residuals, and X is a $k \times p$ design matrix including ones in the first column. Because the effect sizes are assumed to be independent, the covariance matrix of the residuals V_{e} is a diagonal matrix, that is, $V_{\text{e}} = \text{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2]$.

The vector of the estimated regression coefficients via WLS is

$$\hat{\boldsymbol{\beta}} = (X^{\text{T}} V_{\text{e}}^{-1} X)^{-1} X^{\text{T}} V_{\text{e}}^{-1} \mathbf{y}, \tag{7}$$

where $V_{\text{e}}^{-1} = \text{diag}[1/\sigma_1^2, 1/\sigma_2^2, \ldots, 1/\sigma_k^2]$ and the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\hat{V}_{\beta} = (X^{\text{T}} V_{\text{e}}^{-1} X)^{-1}. \tag{8}$$

It is useful to test whether all of the p regression coefficients including the intercept are statistically significant. We may compute a large-sample test statistic

$$\tilde{Q} = \hat{\boldsymbol{\beta}}^{\text{T}} \hat{V}_{\beta}^{-1} \hat{\boldsymbol{\beta}}. \tag{9}$$

Under the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$, the test statistic $\tilde{Q}$ is approximately distributed as a chi-square variate with p degrees of freedom. The above test can easily be modified to test hypotheses on some of the elements in $\boldsymbol{\beta}$, for instance, the regression coefficients excluding the intercept (Hedges & Olkin, 1985).

After testing the significance of all (or some) regression coefficients, we may want to construct an approximate Z test to test the significance of an individual regression coefficient under $H_0 : \beta_i = 0$,

$$Z_2 = \hat{\beta}_i / \sqrt{(\hat{V}_{\beta})_{ii}}, \tag{10}$$

where $(\hat{V}_{\beta})_{ii}$ is the sampling variance of $\hat{\beta}_i$.

Random-Effects Models

Fixed-effects models assume that the population effect sizes share a common value. Many researchers have argued that studies are not direct replications of each other. It is expected that there will be differences in the population effect sizes due to differences with the samples and methods used across studies. Thus, random-effects models should be more appropriate (e.g., Hedges & Vevea, 1998; Hunter & Schmidt, 2000; National Research Council, 1992).

Besides the sampling error, random-effects models include variations in the population effect sizes. According to Hedges and Vevea (1998), the most important issue in determining a fixed-effects versus a random-effects meta-analysis is the nature of the inferences desired. Inferences based on fixed-effects models can be applied only to those studies that have been included in the analysis, whereas inferences based on random-effects models can be generalized beyond the studies in the analysis. Researchers should consider which model is more appropriate for their research questions.

Models without any covariate. The random-effects model is

$$y_i = \beta_{\text{Random}} + u_i + e_i, \tag{11}$$

where β_{Random}, u_i and e_i are the "super" population effect size, the study specific effect, and the sampling error in the *i*th study, respectively. In fixed-effects models, there is only one source of variation, the sampling variance σ_i^2. In contrast, there are two sources of variation in a random-effects model—the sampling variance and the between-studies variance component, $\tau^2 = \text{var}(u_i)$. In the meta-analytic literature, σ_i^2 and $(\tau^2 + \sigma_i^2)$ are known as the conditional and the unconditional variance, respectively.

One common estimator of τ^2, which is based on the quadratic form of the Q statistic in Equation 5, was proposed by DerSimonian and Laird (1986). Their estimator is

$$\hat{\tau}_{DL}^2 = \frac{Q - (k - 1)}{c}, \tag{12}$$

where Q is the statistic of the homogeneity test, k stands for the number of studies, and $c = \sum_{i=1}^{k} w_i - (\sum_{i=1}^{k} w_i^2)/(\sum_{i=1}^{k} w_i)$. When the estimated value is negative, it is truncated to zero. Hedges (1983) also proposed an estimator of τ^2 based on the unweighted method of moments. Maximum likelihood (ML) and restricted maximum likelihood (REML) estimations may also be used in estimating τ^2 (see Viechtbauer, 2005, for an empirical comparison of these estimators).

Once the variance component is estimated, the weighted effect size $\hat{\beta}_{\text{Random}}$ under the random-effects model is

$$\hat{\beta}_{\text{Random}} = \frac{\sum_{i=1}^{k} \tilde{w}_i y_i}{\sum_{i=1}^{k} \tilde{w}_i}, \tag{13}$$

where $\tilde{w}_i = 1/(\sigma_i^2 + \hat{\tau}^2)$ is the weight. The sampling variance $\hat{s}_{\text{Random}}^2$ of $\hat{\beta}_{\text{Random}}$ is estimated by

$$\hat{s}_{\text{Random}}^2 = 1/\sum_{i=1}^{k} \tilde{w}_i. \tag{14}$$

As $\hat{\tau}^2$ is always nonnegative, $\hat{s}_{\text{Random}}^2$ is always larger than $\hat{s}_{\text{Fixed}}^2$ unless $\hat{\tau}^2$ is zero.

Models with covariates. It is sometimes of theoretical interest to include study-specific covariates in random-effects models (e.g., Overton, 1998). These are generally known as mixed-effects models, and they are also widely known as meta-regression in medical research (Berkey, Hoaglin, Mostellar, & Colditz, 1995; Thompson & Higgins, 2002). Mixed-effects models include both fixed and random effects. The fixed effects are the regression coefficients due to the study-specific covariates, whereas the random effects are the unexplained study-specific effects after controlling for the covariates. The model in matrix notation is

$$\mathbf{y} = X\boldsymbol{\beta} + I_k\mathbf{u} + \mathbf{e}, \tag{15}$$

where $\mathbf{y}$ is a $k \times 1$ vector of effect sizes, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed-effects regression coefficients including the intercept, $\mathbf{u}$ is a $k \times 1$ vector of study-specific random effects with $\mathbf{u} \sim N(\mathbf{0}, I_k\tau^2)$, $\mathbf{e}$ is a $k \times 1$ vector of residuals, X is a $k \times p$ design matrix that includes ones in the first column, and I_k is a $k \times k$ identity matrix. Because the effect sizes are assumed to be independent, the conditional covariance matrix of the residuals V_e is a diagonal matrix, that is, $V_e = \text{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2]$.

Raudenbush (1994) proposed a method of moments estimator on τ^2 under the mixed-effects meta-analysis. When $\hat{\tau}^2$ is available, a WLS with Equations 7 and 8 can be used to obtain the parameter estimates and their asymptotic covariance matrix by using a new weight, $\tilde{w}_i = 1/(\sigma_i^2 + \hat{\tau}^2)$. Besides using the method of moments, multilevel models may also be used to analyze random- and mixed-effects meta-analyses (e.g., Hox, 2002; Hox & de Leeuw, 2003; Konstantopoulos & Hedges, 2004; Raudenbush, 1994; Raudenbush & Bryk, 2002). From a mathematical point of view, the model in Equation 15 is the most general model of all of the models that have been mentioned thus far. By dropping the term $I_k\mathbf{u}$, fixed-effects meta-analyses with or without covariates are special cases of Equation 15.

Transforming Data to Achieve Identically Distributed Errors

Fixed-effects models. A WLS regression is frequently used to analyze a fixed-effects meta-analysis by using $w_i = 1/\sigma_i^2$ as the weight (e.g., Hedges & Olkin, 1985). The WLS criterion F_{WLS} to be minimized in order to obtain the parameter estimates is

$$F_{\text{WLS}} = (\mathbf{y} - X\boldsymbol{\beta})^{\text{T}} V_e^{-1} (\mathbf{y} - X\boldsymbol{\beta}). \tag{16}$$

An alternative parameterization is to transform all variables including the intercept by $W^{1/2} = \text{diag}[1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_k]$ (e.g., Kalaian & Raudenbush, 1996; Konstantopoulos, 2008; Kutner, Nachtsheim, Neter, & Li, 2005; Raudenbush, Becker, & Kalaian, 1988). After the transformation, the fixed-effects model becomes

$$W^{1/2}\mathbf{y} = W^{1/2}X\boldsymbol{\beta} + W^{1/2}\mathbf{e}$$

$$y^* = X^*\boldsymbol{\beta} + \mathbf{e}^*, \quad (17)$$

where $\mathbf{y}^* = W^{1/2}\mathbf{y}$, $X^* = W^{1/2}X$, and $\mathbf{e}^* = W^{1/2}\mathbf{e}$. It can be readily shown that $\mathbf{e}^*$ is distributed with a known identity matrix I_k by considering

$$\text{var}(\mathbf{e}^*) = W^{1/2}\text{var}(\mathbf{e})W^{1/2}$$

$$W^{1/2}V_eW^{1/2} = I_k, \quad (18)$$

where $W = V_e^{-1}$.

Because the transformed error $\mathbf{e}^*$ is assumed to be independent and identically distributed with a unit variance, ordinary least squares (OLS) can be directly applied to the data. The OLS criterion F_{OLS} to be minimized in order to obtain the parameter estimates is

$$F_{\text{OLS}} = (\mathbf{y}^* - X^*\boldsymbol{\beta})^{\text{T}}(\mathbf{y}^* - X^*\boldsymbol{\beta}). \quad (19)$$

The parameter estimates based on an OLS regression are equivalent to those based on Equation 7 in a meta-analysis, whereas the standard errors reported by the OLS regression have to be adjusted by a factor.[3]

Random-effects models. The above transformation may also be applied to random- and mixed-effects models. The mixed-effects model based on the transformed data is

$$W^{1/2}\mathbf{y} = W^{1/2}X\boldsymbol{\beta} + W^{1/2}I_k\mathbf{u} + W^{1/2}\mathbf{e}$$

$$\mathbf{y}^* = X^*\boldsymbol{\beta} + I_k^*\mathbf{u} + \mathbf{e}^*, \quad (20)$$

where $I_k^* = W^{1/2}I_k$. After the transformation, $\mathbf{e}^*$ is assumed to be distributed with a known identity matrix I_k. It should be noted that the same transformation with $W^{1/2}$ is applied regardless of whether the model is a fixed- or random-effects one because the conditional variance σ_i^2 is the same under both fixed- and random-effects models.

Mathematically, under fixed-effects models a meta-analysis using WLS with weights and one using an OLS regression on the transformed data are equivalent. For random-effects models, the transformed errors are assumed to be conditionally identically distributed with a unit variance. Transformation allows us to have effect sizes that are assumed to be conditionally identically distributed with a unit variance and to exclude the weights from the analysis. In other words, studies in a meta-analysis can be treated as subjects in a structural equation model. This transformation is the basis for the following SEM approach.

A Structural Equation Modeling Approach

Structural equation modeling (SEM), also known as a covariance structure analysis or a correlation structure analysis, is usually used to fit hypothetical models on the first and second moments (mean vector, covariance, or correlation matrices; e.g., Bentler, 2004; Bollen, 1989; Jöreskog & Sörbom, 1996; L. K. Muthén & Muthén, 2007; Neale, Boker, Xie, & Maes, 2006). Recently, it has been extended to the direct analysis of raw data with missing data (e.g., Arbuckle, 1996; Neale, 2000). Chi-square statistics and goodness-of-fit indices may be used to determine the overall model fit of the proposed model, whereas the significance of an individual parameter estimate can be tested by the parameter estimate divided by its standard error, which has an asymptotic normal distribution (Bollen, 1989).

To illustrate the equivalence between the conventional meta-analysis and the SEM-based meta-analysis, a summary of 20 simulated studies reported in Hox (2002; Hox & de Leeuw, 2003) is listed in Table 1. Table 1 shows the effect size (in Hedges's *d*), its sampling variance, and one continuous study characteristic—the duration of the experimental intervention in terms of weeks (see Hox, 2002, for details). The variable $W^{1/2}$ used for the transformation and the transformed variables are also listed in Table 1.

The computer program MiMa (Viechtbauer, 2006), implemented in R (R Development Core Team, 2008), was used to conduct the conventional meta-analysis reported here, whereas Mplus (Version 5; L. K. Muthén & Muthén, 2007) was used to conduct the SEM-based meta-

[3] When common statistical packages such as SPSS and SAS are used to conduct a meta-analysis with a WLS regression or with an OLS regression on the transformed data, the reported standard errors are incorrect. They have to be adjusted by $\widehat{SE}(\hat{\beta}_j)_{\text{correct}} = \widehat{SE}(\hat{\beta}_j)_{\text{reported}}/\sqrt{MS_e}$, where $\widehat{SE}(\hat{\beta}_j)_{\text{reported}}$ is the standard error on $\hat{\beta}_j$ reported by the statistical packages and MS_e is the mean square error (see Hedges & Olkin, 1985; Konstantopoulos & Hedges, 2004). The reason for this is that the error is assumed to be distributed with a known variance in a meta-analysis. However, the error variance is still estimated in a WLS or OLS regression. Thus, it is necessary to adjust the $\widehat{SEs}$ by rescaling the error variance to a fixed value.

Table 1
Simulated Data Set From 20 Studies in Hox (2002)

Study	d	var(d)	Weeks	$W^{1/2}$	d[a]	Intercept[a]	Weeks[a]	Two-class solution	Three–four-class solutions
1	−0.264	0.086	3	3.4100	−0.9002	3.4100	10.2299	1	1
2	−0.230	0.106	1#	3.0715	−0.7064	3.0715	3.0715#	1	1
3	0.166	0.055	2	4.2640	0.7078	4.2640	8.5280	1	1
4	0.173	0.084	4	3.4503	0.5969	3.4503	13.8013	1	1
5	0.225	0.071	3#	3.7529	0.8444	3.7529	11.2588#	1	1
6	0.291	0.078	6	3.5806	1.0419	3.5806	21.4834	1	1
7	0.309	0.051	7	4.4281	1.3683	4.4281	30.9965	1	1
8	0.435	0.093	9	3.2791	1.4264	3.2791	29.5122	1	2
9	0.476	0.149	3#	2.5906	1.2331	2.5906	7.7719#	1	2
10	0.617	0.095	6	3.2444	2.0018	3.2444	19.4666	1	2
11	0.651	0.110	6	3.0151	1.9628	3.0151	18.0907	1	2
12	0.718	0.054	7	4.3033	3.0898	4.3033	30.1232	1	2
13	0.740	0.081	9#	3.5136	2.6001	3.5136	31.6228#	1	2
14	0.745	0.084	5	3.4503	2.5705	3.4503	17.2516	1	2
15	0.758	0.087	6	3.3903	2.5699	3.3903	20.3419	1	2
16	0.922	0.103	5	3.1159	2.8728	3.1159	15.5794	1	2
17	0.938	0.113	5#	2.9748	2.7904	2.9748	14.8741#	1	2
18	0.962	0.083	7	3.4711	3.3392	3.4711	24.2974	1	2
19	1.522	0.100	9	3.1623	4.8130	3.1623	28.4605	2	3
20	1.844	0.141	9	2.6631	4.9108	2.6631	23.9681	2	3

Note. The data were obtained from Table 8.2 of Hox (2002, p. 146). Data points followed by the pound (#) sign were randomly deleted in the illustration on handling missing covariates.
[a] Data transformed by multiplied by $W^{1/2}$.

analyses. The Mplus codes for the analyses are included in the Appendix.[4]

Fixed-Effects Models

Models without any covariate. A model without any covariate was first analyzed using a conventional meta-analysis. The Q statistic was 49.585 ($df = 19$), $p < .001$. The weighted effect size (and its $\widehat{SE}$) was 0.550 (0.065). The z statistic was 8.465, which was statistically significant at $\alpha = .001$. To fit a fixed-effects meta-analysis in SEM, the model in Equation 17 was fitted with the transformed intercept as the design matrix, that is, $X_0^* = W^{1/2}\mathbf{1}$, where $\mathbf{1}$ is a vector of ones.

Figure 1 shows the graphical model. Using conventional SEM notation, squares, circles, and triangles represent the observed variables, the latent variables, and the means, respectively. There are two points that require attention. First, instead of estimating the error variance on $\mathbf{y}^*$, it is fixed as 1. The $\widehat{SEs}$ reported in the SEM-based meta-analysis are now correct. There is no need to make any corrections of the sort mentioned in footnote 3. Second, the intercept of $\mathbf{y}^*$ is fixed as 0 because the weighted effect size is now represented by $\hat{\beta}_0$. This constraint is crucial when applying the SEM-based meta-analysis. The $\hat{\beta}_0$ (and its $\widehat{SE}$) was 0.550 (0.065). The z statistic was 8.465. The results based on the conventional meta-analysis and the SEM-based meta-analysis are identical. Calculations of the homogeneity test are discussed in a later section.

Model with a covariate. In this analysis, *weeks* was used as a continuous covariate to model the variability in the effect sizes. The design matrix in Equation 6 includes an intercept and a continuous variable. By conducting a conventional meta-analysis, the estimated intercept and regression coefficient (and their $\widehat{SEs}$) were −0.204 (0.170) and 0.135 (0.028), respectively. The z statistics for the intercept and the regression coefficient were −1.205 and 4.816, respectively. Thus, weeks is significant in explaining part of the variation in the effect sizes.

To conduct the same meta-analysis in SEM, a model on the transformed data with two predictors was fitted. Figure 2 shows the model where $X_0^* = W^{1/2}\mathbf{1}$ and $X_1^* = W^{1/2}\textit{weeks}$ were the transformed intercept and the trans-

[4] Other SEM packages such as EQS (Bentler, 2004) and LISREL (Jöreskog & Sörbom, 1996) may also be used to conduct a fixed-effects meta-analysis. As a random slope analysis is required for a random- and mixed-effects meta-analysis, Mplus and Mx are required for these analyses. The complete data set, the Mplus code, and the output are available at my website (http://courses.nus.edu.sg/course/psycwlm/internet/).

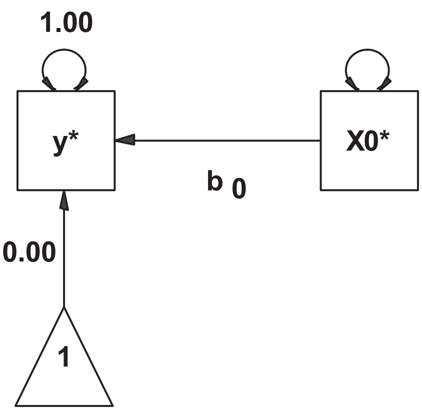

Figure 1. Structural equation model for a fixed-effects meta-analysis without covariate. Squares and triangles represent the observed variables and the means, respectively. **y**$*$, X_0^*, and b_0 are the transformed effect size, the transformed vector of ones, and the weighted effect size under the fixed-effects model, respectively.

formed weeks, respectively. The error variance and the intercept of **y*** were fixed as 1 and 0, respectively, whereas the $\hat{\beta}_0$ and $\hat{\beta}_1$ (and their $\widehat{SEs}$) were -0.204 (0.170) and 0.135 (0.028), respectively. The z statistics for the intercept and the regression coefficient were -1.205 and 4.816. The results based on the conventional meta-analysis and the SEM-based meta-analysis are the same.

Random-Effects Models

Models without any covariate. Because ML is usually used as the estimation method in SEM, ML was used to estimate the variance component in order to compare the results of the conventional meta-analysis and the SEM-based meta-analysis. Issues relating to the use of different estimation methods in estimating the variance component are discussed later. A random-effects model with an intercept model was fitted. The ML estimate of τ^2 reported in MiMa was 0.135, whereas the weighted effect size (and its $\widehat{SE}$) was 0.579 (0.105). The z statistic was 5.520. Thus, the weighted effect size is statistically significant under the random-effects model.

A random-effects meta-analysis can be formulated as a single-level analysis with random slopes in SEM. The model without any covariate in Equation 20 can be expressed as

$$\mathbf{y}^* = I_k^* \mathbf{u} + \mathbf{e}^*, \tag{21}$$

where $\mathbf{u} \sim N(\beta_0 \mathbf{1},\ I_k \tau^2)$. Figure 3 shows the graphical model in which $X_0^* = I_k^*$ is the transformed vector of ones. In Figure 3, u with a dot in the arrow from X_0^* to $\mathbf{y}^*$ represents a random slope that varies across studies. Thus, u_i in the ith study is treated as a random variable. It can be easily shown that the mean of u_i is the weighted effect size $\hat{\beta}_0$, whereas the variance of u_i (m in Figure 3) is the variance component $\hat{\tau}^2$. Because u_i varies across subjects (studies in the context of a meta-analysis), it is necessary to conduct a random slope analysis (Mehta & Neale, 2005; B. Muthén & Asparouhov, 2002, 2003; L. K. Muthén & Muthén, 2007). The $\hat{\beta}_0$ (and its $\widehat{SE}$) was 0.579 (0.107). The z statistic was 5.406. The $\hat{\tau}^2$ was 0.132, which is comparable to the estimates based on MiMa.

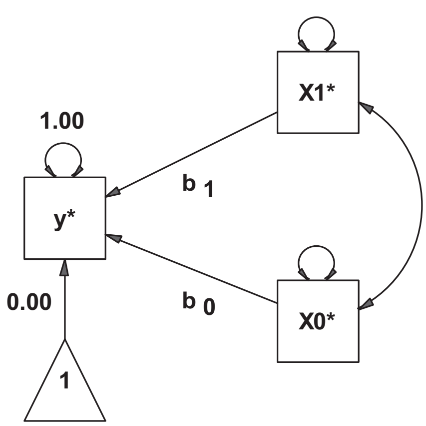

Figure 2. Structural equation model for a fixed-effects meta-analysis with a covariate. Squares and triangles represent the observed variables and the means, respectively. **y**$*$, X_0^*, and X_1^* are the transformed effect size, the transformed vector of ones, and the transformed covariate weeks, respectively. b_0 and b_1 are the estimated intercept and the estimated slope under the fixed-effects model, respectively.

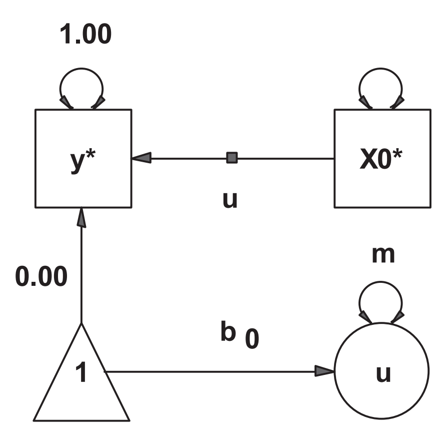

Figure 3. Structural equation model for a random-effects meta-analysis without covariate. Squares, circles, and triangles represent the observed variables, the latent variables, and the means, respectively. **y**$*$ and X_0^* are the transformed effect size and the transformed vector of ones, respectively. u, b_0, and m are the study-specific effect, the weighted effect size, and the variance component under the random-effects model, respectively.

Models with a covariate. The covariate weeks was then added to the model. The $\hat{\tau}^2$ dropped to 0.023 with $\chi^2(18) = 26.392$, $p = .091$ reported by MiMa. The estimated regression coefficients (and their $\widehat{SEs}$) for the intercept and weeks were −0.214 (0.193) and 0.139 (0.032), respectively. The z statistics were −1.109 and 4.352. Thus, the covariate weeks is still significant under the mixed-effects model.

The model in Equation 21 can easily be extended to include covariates in the SEM-based meta-analysis. The model with one covariate is

$$\mathbf{y}^* = X_0^* u + X_1^* \beta_1 + \mathbf{e}^*, \tag{22}$$

where β_1 is the regression coefficient of weeks, $X_0^* = I_k^*$ is the transformed vector of ones, and $X_1^* = W^{1/2}weeks$ is the transformed weeks that does not include the intercept. It should be noted that the intercept β_0 does not appear in Equation 22 because it is absorbed by the mean of u. Figure 4 shows the graphical model with a covariate.

The $\hat{\tau}^2$ was 0.023 in the SEM-based meta-analysis. The mean (and its $\widehat{SE}$) of u was −0.214 (0.171). The regression coefficient (and its $\widehat{SE}$) for weeks was 0.139 (0.036). There are some minor differences in the estimated standard errors between MiMa and Mplus. It is speculated that the differences are due to the iterative procedures and the convergence criteria used in MiMa and Mplus.

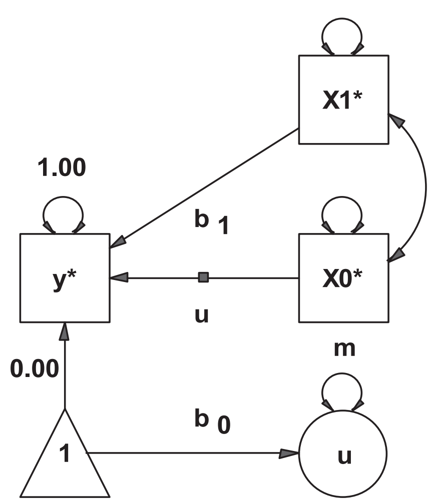

Figure 4. Structural equation model for a mixed-effects meta-analysis with a covariate. Squares, circles, and triangles represent the observed variables, the latent variables, and the means, respectively. $\mathbf{y}*$, X_0^*, and X_1^* are the transformed effect size, the transformed vector of ones, and the transformed covariate weeks, respectively. b_0, b_1, u, and m are the estimated intercept, the estimated slope, the study-specific effect, and the residual variance component under the mixed-effects model, respectively.

Extensions of the SEM-Based Meta-Analysis

In the above examples, I have shown how SEM can be used as a framework for conducting a meta-analysis. Researchers with basic SEM knowledge may find the proposed approach useful. This may stimulate more SEM researchers into applying meta-analytic techniques in their research. However, meta-analysts may wonder whether this new approach brings additional benefits beyond those that can be obtained by carrying out a conventional meta-analysis. In this section I demonstrate how many state-of-the-art techniques implemented in SEM packages can be directly applied to meta-analyses. These include mixture modeling, the bootstrap method, handling missing covariates with the full information maximum likelihood (FIML), and computing new parameters based on other parameters.

Most of these techniques are available in the literature; researchers may implement them to suit their needs. However, implementing their own procedures may not be practical for researchers who lack a programming background. The SEM approach can be applied using the Mplus code set out in the Appendix. It is hoped that SEM-based meta-analysis may provide a favorable alternative to conducting advanced meta-analyses. In addition, SEM-based meta-analysis allows researchers to apply more than one technique in the same meta-analysis. For example, a researcher may want to handle missing covariates and to conduct a mixture model in the same analysis. This may not be a trivial task under a conventional meta-analysis.

Handling Missing Covariates

Study characteristics are usually used as covariates in a meta-analysis. Because the primary studies are conducted by different researchers with different research objectives, it is quite common for different study characteristics to be reported. This could create missing data on covariates of interest to the meta-analyst (Pigott, 1994, 2001). Cooper and Hedges (1994) have called missing data "perhaps the most pervasive practical problem in research synthesis" (p. 525).

Rubin (1987) defined three types of missingness mechanism. The missingness on a variable, say Y, is said to be missing completely at random (MCAR) if the missingness is unrelated to the value of Y itself or to the values of any other variables in the model. This assumption may be rather strong in applied settings. A considerably weaker assumption is missing at random (MAR). MAR means that the missingness on Y is unrelated to the value of Y after controlling for other variables in the analysis. The missingness on Y is said to be not missing at random (NMAR) if the missingness on Y is related to the value of Y itself.

Rubin's (1987) definitions have been directly adopted for meta-analysis (Pigott, 2001; Sutton & Pigott, 2005; Sutton, Abrams, Jones, Sheldon, & Song, 2000). In the context of missing covariates, MCAR means that the missingness of a covariate is unrelated to the value of that covariate or of other variables. MAR means that the missingness of a covariate may be related to other variables, for instance, the effect size or other covariates, whereas NMAR means that the missingness of a covariate is related to the value of that covariate.

As researchers are not very likely to fail to report a study characteristic, for example, mean age of the participants, because of the value of that covariate, Sutton and Pigott (2005) stated that "the assumption that [study-level covariates] are MCAR or MAR may be reasonable and standard missing-data methods may suffice in some situations" (p. 235). It should be noted that the missingness on the effect sizes, for example, a correlation coefficient between job satisfaction and performance, is likely to be NMAR. If the effect sizes are nonsignificant, they are less likely to be reported or published. This is known as publication bias, which is beyond the scope of this article (see Rothstein, Sutton, & Borenstein, 2005).

Pigott (2001) summarized several conventional methods of handling missing covariates in a meta-analysis. These include listwise deletion, mean substitution, and pairwise deletion. However, these methods are generally not recommended (Schafer & Graham, 2002). Pigott suggested using an expectation-maximization (EM) algorithm and multiple imputation (MI) to handle missing covariates. In order to handle missing covariates in a regression analysis, Little (1992) suggested employing MI and ML methods. Schafer and Graham (2002) also recommended using the FIML method, which has been implemented in many SEM packages, and MI to handle missing data with MCAR and MAR.[5]

Assuming conditional multivariate normality in the data, the casewise log-likelihood of the observed data under the FIML method is obtained by maximizing the function

$$\log L_i = K_i - \frac{1}{2}\log\left|\textstyle\sum_i\right| - \frac{1}{2}\,(\mathbf{z_i} - \boldsymbol{\mu}_\mathbf{i})^{\mathrm{T}} \textstyle\sum_i^{-1} (\mathbf{z_i} - \boldsymbol{\mu}_\mathbf{i}), \tag{23}$$

where K_i is a constant that depends on the number of complete data points for the ith case, $\mathbf{z_i}$ is the observed complete vector including the independent and the dependent variables, and $\boldsymbol{\mu}_\mathbf{i}$ and Σ_i are the implied population mean and covariance matrix for the ith case, respectively (e.g., Arbuckle, 1996; Enders, 2001b, 2006; Neale, 2000). The overall log-likelihood of the N cases (studies in the context of meta-analysis) is $\Sigma_{i=1}^{N} \log L_i$. The ML estimates are obtained by maximizing this likelihood function. Because the number of observed data in $\mathbf{z_i}$ is allowed to be different across subjects, missing data can be easily handled.

The evidence from empirical studies supports the contention that FIML performs better than such conventional methods as listwise deletion, pairwise deletion, and mean imputation in handling missing data when the missingness is either MCAR or MAR (see M. W. L. Cheung, 2007b, and Enders, 2001a, for some empirical findings on a comparison of methods of handling missing covariates in the context of a latent growth model and a regression analysis). When the missingness is NMAR, none of the above methods is unbiased (see Schafer, 1997). However, the bias of the FIML is still less than that resulting from listwise deletion, pairwise deletion, and mean substitution (Jamshidian & Bentler, 1999; B. Muthén, Kaplan, & Hollis, 1987).

Because both FIML and MI are asymptotically equivalent (e.g., Graham, Olchowski, & Gilreath, 2007), one may wonder whether the FIML method available in SEM packages is preferable to MI for handling missing covariates. There are two advantages to using FIML. First, the results based on MI are asymptotically equivalent to those based on FIML. More explicitly, they are equal only when the number of imputations in MI approaches infinity. It is generally suggested that three to five imputations are sufficient to obtain excellent results in MI (e.g., Schafer & Olsen, 1998). However, Graham et al. (2007) have recently shown that a few imputations might not be sufficient. They concluded that "in sum, our simulations results show rather clearly that FIML is superior to MI, in terms of power for testing small effect sizes, unless one has a sufficient number of imputations" (p. 212).

Second, SEM assumes conditional normality on the dependent variables conditional on the observed covariates (e.g., Bollen, 1989; B. Muthén, 2004b). The observed covariates do not need to be normally distributed. Dummy and ordinal variables may also be used as observed covariates (see Jöreskog & Sörbom, 1996, for some examples of the utilization of dummy variables in running ANOVA and multivariate analysis of variance [MANOVA] in SEM). When using FIML in SEM, the transformed covariates do not need to be multivariate normal. In contrast, multivariate normality is assumed for all variables in MI (see Schafer, 1997, for a discussion on the effect of nonnormality on MI).

To illustrate the procedures for analyzing missing covariates, 5 out of 20 studies were deleted in the covariate weeks (the data marked with the symbol # in Table 1). The missingness was qualified as MCAR. Three analyses (the listwise deletion analysis, the FIML analysis, and the MI analysis with

[5] Different SEM packages use slightly different algorithms to obtain the ML estimates in the presence of missing data. Amos, LISREL, and Mx directly implement Equation 23, whereas Mplus (B. Muthén, 2004b) and EQS 6 (Bentler, 2004) use an EM algorithm (see Schafer & Graham, 2002, for a discussion on how SEM packages implement the ML methods).

100 imputations) based on the fixed-effects model were conducted.[6] The parameter estimates (and their $\widehat{SEs}$) of the intercept and the weeks for the listwise deletion analysis were −0.313 (0.226) and 0.152 (0.036), respectively. The z statistics were −1.385 and 4.212. The parameter estimates (and their $\widehat{SEs}$) of the intercept and the weeks for the MI analysis were −0.315 (0.200) and 0.149 (0.033), respectively. The z statistics were −1.575 and 4.579. Regarding the FIML analysis, the parameter estimates (and their $\widehat{SEs}$) of the intercept and the weeks were −0.316 (0.208) and 0.150 (0.034), respectively. The z statistics were −1.521 and 4.405.

When compared to the results of the full sample, the parameter estimates and their standard errors based on the sample with missing covariates are smaller (in absolute values) and larger, respectively. This is expected because some information will have been lost due to the missing data. The standard errors based on the MI and the FIML analyses were slightly smaller than those based on the listwise deletion analysis because in an MI or FIML analysis it is not necessary to throw out the whole case in the presence of missing values.

Quantifying Heterogeneity in a Meta-Analysis

Quantifying heterogeneity is essential in a meta-analysis (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). When the estimated variance component is small, the results based on a fixed-effects and a random-effects model are similar. Some researchers may choose to use random-effects models only if the effect sizes are heterogeneous. This approach is called the conditional random-effects model (Hedges & Vevea, 1998). One problem with using the Q statistic to quantify heterogeneity is that the Q statistic depends on the number of studies in a meta-analysis. When the number of studies increases, the Q statistic also increases.

Higgins and Thompson (2002) proposed three indices to quantify heterogeneity. Two of them are described here. The first is the H^2 index,

$$H^2 = \frac{Q}{k-1}. \tag{24}$$

The second is the I^2 index,

$$I^2 = \frac{H^2 - 1}{H^2}. \tag{25}$$

When I^2 is negative, it is truncated to zero. The H^2 and I^2 indices can be interpreted as the relative excess in Q statistic per degree of freedom and the proportion of the total variation due to the heterogeneity between studies, respectively. Higgins and Thompson further put forward the following formula to construct an approximate confidence interval (CI) on $\ln(H)$:

$$\exp(\ln(H) \pm Z_\alpha SE(\ln(H))), \tag{26}$$

where Z_α is the $(1 - \alpha/2)$ quantile of the standard normal distribution and

$$SE(\ln(H)) = \begin{cases} \dfrac{1}{2}\dfrac{\ln(Q) - \ln(k-1)}{\sqrt{2Q} - \sqrt{2k-3}} & \text{if } Q > k \\ \sqrt{\dfrac{1}{2(k-2)}\left(1 - \dfrac{1}{3(k-2)^2}\right)} & \text{if } Q \le k \end{cases}.$$

The CI of the I^2 index may be transformed from the CI of the H^2 index by Equation 25. Because the I^2 index is easier to interpret, it was used in this illustration. The Q statistic for the 20 studies was 49.585 (df = 19), $p < .001$. The calculated I^2 index and its 95% CI were 0.617 and (0.378, 0.764), respectively. Therefore, 62% of the variation of the effect sizes is due to the population heterogeneity, whereas 38% of the variation is due to the sampling error.

So far, in all of the illustrations of the SEM-based meta-analysis, the error variance of **y*** was always fixed at 1. By so doing, the correct $\widehat{SEs}$ can be directly obtained from the output. The error variance of **y*** is set free in estimating the Q statistic. The ML estimate of the error variance of **y*** in the fixed-effects model without any covariate is $\hat{\sigma}^2_{e*} = \sum_{i=1}^{k} (y^*_i - 1^*\hat{\beta}_{\text{Fixed}})^2/k$, where $1^* = w_i^{1/2}*1$, which is equivalent to $\sum_{i=1}^{k} w_i(y_i - \hat{\beta}_{\text{Fixed}})^2/k$.[7] It should be noted that k instead of $(k - 1)$ is used in the ML estimator. Then, it is readily clear that

$$Q = k\hat{\sigma}^2_{e*}. \tag{27}$$

The above equation also applies to testing the residual heterogeneity in the presence of covariates. Once the Q statistic is estimated, it is easy to calculate the H^2 and I^2 indices.

[6] Missing covariates are not allowed in the random-effects models in Mplus because it is not possible to estimate the variance component on the basis of the missing values. Cases with missing covariates will be deleted in the random slope analyses. An alternative approach is to use MI, which is also available in Mplus.

[7] It can be easily shown that $\sum_{i=1}^{k} (y^*_i - 1^*\hat{\beta}_{\text{Fixed}})^2 = \sum_{i=1}^{k} w_i(y_i - \hat{\beta}_{\text{Fixed}})^2$ by considering the following:

$$\sum_{i=1}^{k} w_i(y_i - \hat{\beta}_{\text{Fixed}})^2 = \sum_{i=1}^{k} (w_i^{1/2}y_i - w_i^{1/2}\hat{\beta}_{\text{Fixed}})^2$$

$$= \sum_{i=1}^{k} (y^*_i - 1^*\hat{\beta}_{\text{Fixed}})^2,$$

where $1^* = w_i^{1/2}*1$.

Many SEM packages allow for the calculation of new parameter estimates that are functions of other parameter estimates in a model. Standard errors (based on the delta method), Wald CIs, and bootstrap CIs on the parameter estimates can be directly obtained from the output (see M. W. L. Cheung, 2007a, in press, for some applications of this approach in constructing CIs). By using the SEM-based meta-analysis, these methods for constructing CIs are readily available to meta-analysts.

Using the approaches in M. W. L. Cheung (2007a, in press), the estimated Q statistic and I^2 index (and their $\widehat{SEs}$) in the SEM-based meta-analysis were 49.585 (15.680) and 0.617 (0.121), respectively. A 95% CI on the I^2 index was also constructed. The Wald CI and bias-corrected bootstrap CI with 2,000 replications were obtained in Mplus. On the I^2 index, the 95% Wald and bias-corrected bootstrap CIs were (0.379, 0.854), and (0.263, 0.774), respectively. These CIs are comparable to the one proposed by Higgins and Thompson (2002). Because using SEM to quantify heterogeneity in meta-analysis is a new approach, further research may address which method has the best coverage probabilities and optimal length.

Addressing Heterogeneity With Mixture Models

As researchers tend to accept that population effect sizes are heterogeneous, addressing heterogeneity properly is an important issue in meta-analysis (Thompson & Sharp, 1999). One common approach to handling heterogeneity is the use of random- or mixed-effects models. Random-effects models assume that the population effect sizes are distributed with an unknown variance that needs to be estimated.

An alternative approach to handling heterogeneity is to use finite mixture models (McLachlan & Peel, 2000). Finite mixture models assume that the effect sizes are drawn from several distinct unobserved populations. Given the same population, the effect sizes are homogeneous within that population. They are useful for identifying hidden populations. Several researchers (e.g., Böhning, 1999, 2005; Thomas, 1989; Xia, Weng, Zhang, & Li, 2005) have extended mixture models to handle the heterogeneity of effect sizes in meta-analysis.[8] The basic procedure is to conduct a fixed-effects meta-analysis with covariates. When there is still large residual heterogeneity after controlling for the covariates, mixture models along with the covariates are used to identify the number of the hidden populations. On the basis of the predicted classes, researchers may check whether any consistent patterns are shown in these classes.

Mixture models have been integrated into some SEM packages such as Mplus and Mx (e.g., Lubke & Muthén, 2005; B. Muthén, 2001, 2004a; Neale, 2000; Neale et al., 2006). They have been applied to address unobserved populations in latent growth modeling (B. Muthén, 2001, 2004a) and heterogeneity in factor analysis (Lubke & Muthén, 2005; Yung, 1997). By using the SEM-based meta-analysis, mixture models can be easily accessible to meta-analysts. One advantage of using an SEM package such as Mplus to conduct mixture modeling is that multiple random starting values can be easily generated. Multiple maxima of the likelihood often exist in mixture models (McLachlan & Peel, 2000). Multiple random starting values may be used to minimize the chance of having local maxima.

Recall that the Q statistic and the $\hat{\tau}^2$ in the SEM-based meta-analysis are 49.585 (df = 19) and 0.132, respectively, whereas the estimated variance component after controlling for the covariate weeks is only 0.023, which is nonsignificant. As a demonstration, I did not include the covariate weeks in the mixture analyses; otherwise, one single class would be found after controlling for the covariate weeks. In this illustration, three mixture models based on two, three, and four mixtures were fitted. Figure 5 shows the graphical model where c is a latent categorical variable. The model implies that different classes of effect sizes may have different $\hat{\beta}_0$. Posterior class probabilities may be used to predict the studies into classes (see L. K. Muthén & Muthén, 2007).

The predicted class memberships are reported in Table 1. Regarding the four-class solution, only three classes were estimated. Thus, only two- and three-class solutions were considered. There are several indices that may be used to compare solutions with a different number of mixtures (see McLachlan & Peel, 2000). One common statistic is the Bayesian information criterion (BIC; Schwartz, 1978). The model with the smallest BIC is preferred. The BICs for the two- and three-class solutions were 85.521 and 89.250, respectively. Another statistic is the adjusted likelihood ratio test (aLRT; Lo, Mendell, & Rubin, 2001). Lubke and Muthén (2007) have recently shown that the aLRT performed very well in identifying the number of classes. The aLRT in comparing the two- and three-class solutions was 1.939 (p = .060). This indicates that the three-class solution is not better than the two-class solution. Based on the BIC and aLRT, it seems that the two-class solution is sufficient.

[8] Mixture models have also been extended to general mixed-effects models (e.g., Hall & Wang, 2005; Verbeke & Lesaffre, 1996; Wang, Schumitzky, & D'Argenio, 2007). Because mixed-effects meta-analysis is a special case of general mixed-effects models, it is possible to modify these models for mixed-effects meta-analysis. However, researchers have to consider two practical issues. First, not all mixed-effects packages are capable of fitting mixed-effects meta-analysis because the Level 1 variances are fixed as known values in a mixed-effects meta-analysis (see Hox, 2002, for a discussion). Second, multiple maxima of the likelihood often exist in mixture models (McLachlan & Peel, 2000). Researchers have to generate multiple random starting values in order to minimize the chance of local maxima.

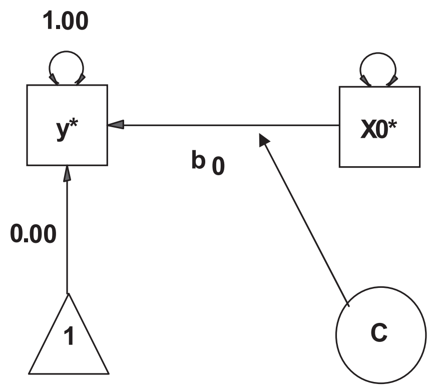

Figure 5. Structural equation model for a fixed-effects meta-analysis with a mixture model. Squares, circles, and triangles represent the observed variables, the latent variables, and the means, respectively. $\mathbf{y}*$, X_0^*, b_0, and c are the transformed effect size, the transformed vector of ones, the weighted effect size, and the latent categorical variable under the mixture model, respectively. It should be noted that the weighted effect sizes can be different in different latent classes.

The weighted effect sizes (and their $\widehat{SEs}$) for the two-class solution were 0.463 (0.084) and 1.602 (0.161). These estimates would be slightly different from those obtained by running two separate analyses on the predicted classes. The reason for this is that group membership is treated as a latent variable in a mixture model, whereas it is treated as fixed and known in running two separate analyses. By checking the two-class solution, Studies 19 and 20 were classified as belonging to a different class from other studies. The Q statistics and their p values for these two classes were 26.6 ($df = 17$, $p = .06$) and 0.4 ($df = 1$, $p = .52$), respectively. This also supports the argument that the two-class solution is good enough to account for the heterogeneity when no covariate is included.

The data in Table 1 were actually generated from a population model $y_i = 0.15*Weeks_i + e_i$ with a mean effect size of 0.6 across all studies (see Hox & de Leeuw, 2003). Conceptually speaking, each level of the covariate weeks can be considered as a class, and the classes are separated by 0.15 units. Some may expect to find as many classes as the number of levels in the covariate weeks, that is, eight classes in this example. However, the number of classes identified in a mixture model depends on the class separation and the within-class sample size (Lubke & Neale, 2006). In this particular example, the class separation (0.15 units in each class) is not large enough and the number of studies within a particular level is also quite small. Thus, only two classes were identified.

Lubke and Muthén (2007) also investigated the performance of the factor mixture models via a computer simulation. One general finding was that the parameter estimates are still very good even for a small degree of separation among the classes, whereas the correctness of class predictions depends on the degree of separation among the classes. This means that the estimated regression coefficients were very good. However, the predicted class membership may be inaccurate if the population effect sizes are similar among the classes. Because applications of mixture models in meta-analysis are relatively new, further studies may be needed to address the empirical performance of this approach.

Discussion and Future Directions

The present article proposes a model to integrate meta-analysis into the SEM framework. By applying an appropriate transformation to the data, studies in a meta-analysis can be treated as subjects in a structural equation model. The illustrations demonstrated the equivalence between conventional meta-analysis and SEM-based meta-analysis. Many state-of-the-art techniques available in SEM can be directly applied to meta-analysis. Integrating meta-analysis into SEM opens several new research directions for methodological development. The following is a brief discussion of some possible directions for future research, as well as related issues that will require further exploration.

Model Assessment With Chi-Square Test Statistics and Goodness-of-Fit Indices

Thus far, the homogeneity of the effect sizes was tested based on the conventional Q statistic in meta-analysis (Equations 5 and 27). The likelihood ratio (LR) or chi-square statistic is the fundamental test statistic in SEM. An LR statistic may also be derived to test the homogeneity of effect sizes. Let us consider a fixed-effects model without any covariates. The model for the transformed data in an SEM-based meta-analysis can be expressed as

$$\mathbf{y}* = X*\beta_{\text{Fixed}} + \mathbf{e}*, \tag{28}$$

where $X* = W^{1/2}\mathbf{1}$ and β_{Fixed} is a population effect size. Under the null hypothesis that all effect sizes are equal, the error variance of $\mathbf{e}*$ is exactly 1. In the SEM-based meta-analysis, this model is fitted by fixing the error variance of $\mathbf{e}*$ at 1. It is labeled as Model 0 here and its associated chi-square statistic is $\chi_0^2(2)$.

If the null hypothesis is incorrect, the estimated variance of $\mathbf{e}*$ in Equation 28 will be larger than 1. This is because the study-specific effect is combined with the error variance. This model may be fitted in SEM-based meta-analysis by freeing the estimated variance of $\mathbf{e}*$. It is labeled as Model 1 here with its associated chi-square statistic $\chi_1^2(1)$.

Under the null hypothesis $H_0 : \sigma_{e*}^2 = 1$, a chi-square difference test may be used to compare these two nested models: $\Delta\chi^2 = \chi_0^2 - \chi_1^2$ with $\Delta df = 1$. Because the null hypothesis $H_0 : \sigma_{e*}^2 = 1$ is tested on a boundary condition, we have to adjust the critical value in order to conduct a better test. If the observed p value is larger than 2α, we reject the null hypothesis at the α level.[9] Regarding our example in Table 1, the chi-square statistics for Models 0 and 1 are $\chi_0^2(2) = 14.453$ and $\chi_1^2(1) = 3.027$, respectively. The chi-square difference test is $\Delta\chi^2(1) = 11.426$, $p < .001$. Recall that the Q statistic for this example is 49.585 ($df = 19$), $p < .001$.

Although the above LR statistic may also be used to test the homogeneity of effect sizes, the Q statistic is preferred to the LR statistic in the context of meta-analysis. Viechtbauer (2007b) compared several statistics in testing the population heterogeneity. These included the Q statistic, the LR tests based on ML and REML, and the Wald and score tests. Generally, the Q statistic was found to have the best control over the Type I error rate and to have reasonable statistical power.

One of the reasons for the above observation may be due to the fact that the Q statistic has an exact chi-square distribution, even for two studies where the within-study sample sizes are sufficiently large. However, the LR statistic must rely on both a large within-study sample size and a large number of studies. Furthermore, large sample sizes (number of studies) are required for the valid application of LR statistics. This is clear when integrating meta-analysis into the SEM framework; studies in a meta-analysis become subjects in an SEM analysis.

Besides formal chi-square statistics, many goodness-of-fit indices are available in SEM. Much research has been devoted to the issue of how to assess the model fit using goodness-of-fit indices (see Bollen & Long, 1993). Many goodness-of-fit indices such as the comparative fit index (CFI) and the nonnormed fit index (NNFI) use an independence model (a model in which all of the variables are uncorrelated) as the reference model for comparison. A null model on the transformed data is not interpretable in a meta-analysis. Thus, many goodness-of-fit indices may not be well defined in the context of meta-analysis. Because using SEM to conduct a meta-analysis is a new approach, future research is definitely needed to explore whether or not the rich knowledge on model fitting techniques in SEM is applicable to meta-analysis.

Methods of Estimating the Variance Component

Method of moments is usually used to estimate the variance component (e.g., DerSimonian & Laird, 1986; Hedges & Vevea, 1998). There are several appealing reasons for using this method. First, close-form formulas are usually available; no iterative procedure is required. Second, it is distribution free. The first and the second moments are used to estimate the variance component and its standard error, respectively (see Biggerstaff & Tweedie, 1997, for the details).

Maximum likelihood (ML or REML) methods are the preferred choices when more complicated meta-analyses are involved. For example, ML methods are used to handle missing data and to handle heterogeneity with mixture models (Böhning, 1999, 2005). It is well known that the estimated variance component in ML is negatively biased, especially in small samples, whereas REML may be used to correct the bias. Because the number of studies in a meta-analysis may be relatively small, meta-analysts should be cautious about applying the SEM-based meta-analysis if their research interest is to estimate and compare the variance components of the random effects.

The REML estimate of the variance component can be approximated by multiplying the ML estimate by $k/(k - p)$, where k is the number of studies and p is the number of fixed effects (p is 1 in the random-effects model and p is 1 plus the number of covariates in the mixed-effects models; see Overton, 1998; Viechtbauer, 2005). This approximation can be easily obtained in the SEM approach, as many SEM packages allow us to compute functions of other parameters.

It is also possible to use the variance component estimate based on the method of moments (or the REML) in the SEM-based meta-analysis. Once the estimated variance component (based on the REML or the method of moments) is available, it may be used to fix the variance of u (m in Figures 3 and 4) in the SEM-based meta-analysis. However, it should be remembered that this is just an approximation. Further research may be necessary to address how different estimation methods can be implemented in the SEM-based meta-analysis.

Robust Standard Errors on the Variance Component

When the research questions are directly related to the variance component, meta-analysts are interested in testing the significance of the variance component (Viechtbauer,

[9] It may not be obvious that the null hypothesis $H_0 : \sigma^2_{e*} = 1$ is tested at a boundary. The H_0: $\tau^2 = 0$ in a conventional meta-analysis is equivalent to $H_0 : \sigma^2_{e*} = 1$ on the transformed data in an SEM-based meta-analysis, whereas H_1: $\tau^2 > 0$ in a conventional meta-analysis is translated into $H_1 : \sigma^2_{e*} > 1$ in an SEM-based meta-analysis. Because τ^2 cannot be negative, σ^2_{e*} cannot be smaller than 1, theoretically. Because of the boundary condition, the $\Delta\chi^2$ is distributed as a 50:50 mixture of a degenerate random variable with all of its probability mass concentrated at zero and a chi-square random variable with 1 degree of freedom. One simple strategy to correct for this issue is to use 2α instead of α as the alpha level. Thus, we should reject the null hypothesis when the observed p value is larger than .10 for $\alpha = .05$. Readers may refer to Viechtbauer (2007b) and Stoel, Galindo-Garre, Dolan, and van den Wittenboer (2006) for a discussion on this issue in the context of meta-analysis and SEM. The estimates of τ^2 can sometimes be negative because of the sampling fluctuation. As there is a 1:1 mapping between τ^2 in a meta-analysis and σ^2_{e*} in an SEM-based meta-analysis, the estimates of σ^2_{e*} can also be smaller than 1 in some cases.

2007b) or in constructing approximate CIs on the variance component (Viechtbauer, 2007a). Although estimating the variance component with the method of moments does not assume normality of the effect sizes, testing and constructing CIs on the variance component requires such an assumption. Moreover, the testing and constructing of CIs on the weighted effect size with the standard error estimated by Equation 14 also rely on the assumption of normality on the part of the effect sizes. In other words, most methods, including the method of moments, ML, and REML, implicitly or explicitly assume normality of the effect sizes or of the variance component.

The assumption of the normality of the effect sizes has been called into question. Several robust procedures have been suggested in meta-analysis (e.g., Demidenko, 2004; Sidik & Jonkman, 2006). One approach is to use a parametric and nonparametric bootstrap (e.g., Adams, Gurevitch, & Rosenberg, 1997; Takkouche, Cadarso-Suarez, & Spiegelman, 1999; Van den Noortgate & Onghena, 2005). For the parametric bootstrap, the parameter estimates are used as population values to generate the CIs of interest while the data are resampled with replacement to generate the CIs of interest in the nonparametric bootstrap (see Davison & Hinkley, 1997).

The robustness of test statistics and standard errors against nonnormality is also an important topic of research in SEM (e.g., Yuan & Bentler, 2007). Most SEM packages have implemented some form of robust statistics that behaves better under nonnormality. As meta-analysis becomes integrated in SEM, it is of interest to see how well these robust standard errors work in the context of meta-analysis. Moreover, many SEM packages have functions to conduct bootstrapping. It is also of practical interest to explore the usefulness of the bootstrap technique in meta-analysis.

A Final Remark on Modeling Meta-Analytic Data in SEM

This article presents a SEM-based approach to conducting meta-analysis. Readers should be reminded that data in meta-analyses are usually more complex in terms of conceptualization and data collection than data in primary studies. Inexperienced readers may wrongly conclude that once the appropriate transformation described in this article has been made and the study results in a meta-analysis have become subjects in a structural equation model, there is no difference between applying SEM to primary data versus meta-analytic data in any phase of the research process.[10]

Similar concerns have also been raised about the usefulness of path diagrams in SEM and multilevel models (see Curran & Bauer, 2007). Many SEM packages, for example, AMOS, EQS, Mx, and LISREL, allow the use of path diagrams as input. Inexperienced users may fit ill-formulated models because fitting a model is similar to drawing a path diagram. Curran and Bauer (2007) "firmly believe[d] that the advantages of model diagrams vastly outweigh their potential weakness" (p. 296). I also strongly believe that the advantages of using the SEM approach to conducting a meta-analysis also outweigh its potential misuse by inexperienced researchers. This article highlights the fact that applied researchers can access several advanced statistical techniques in meta-analysis via some popular SEM packages. Instead of spending time in implementing these algorithms, applied researchers may now focus on the proper applications of these techniques.

Finally, I have to emphasize that a meta-analysis is more than just a statistical analysis. In this article, I have mainly focused on the statistical methods of meta-analysis. There are, however, other design and conceptual issues that are specific to meta-analysis. For example, studies are usually not direct replicas of each other (e.g., Hunter & Schmidt, 2000). It is sometimes difficult to define what a population is. It is also generally agreed that the published data are usually biased against nonsignificant findings (see Rothstein, Sutton, & Borenstein, 2005). Special care is required to handle meta-analytic data. Cooper and Hedges (1994) and Lipsey and Wilson (2001) provided a complete treatment of issues related to meta-analysis. Readers are strongly advised to consult their works before conducting a meta-analysis.

To summarize, this article provides a start to integrating meta-analysis into the SEM framework. It is hoped that this new line of research will benefit both meta-analysts and researchers in SEM. As many SEM packages are becoming more and more powerful, some of these new developments will prove useful to meta-analysts. A unified framework may one day be available to researchers who conduct meta-analysis, SEM, and other types of statistical analysis.

[10] I thank an associate editor for bringing this issue to my attention.

References

Adams, D. C., Gurevitch, J., & Rosenberg, M. S. (1997). Resampling tests for meta-analysis of ecological data. *Ecology, 78,* 1277–1283.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28,* 135–167.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41,* 257–278.

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, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499–525). San Diego, CA: Academic Press.

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: Russell Sage Foundation.

Bentler, P. M. (2004). *EQS* 6 [Computer software]. Encino, CA: Multivariate Software.

Berkey, C. S., Hoaglin, D. C., Mostellar, F., & Colditz, G. A. (1995). A random effects regression model for meta-analysis. *Statistics in Medicine, 14,* 395–411.

Biggerstaff, B. J., & Tweedie, R. L. (1997). Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine, 16,* 753–768.

Böhning, D. (1999). *Computer assisted analysis of mixtures and applications: Meta-analysis, disease mapping, and others.* London: Chapman & Hall/CRC.

Böhning, D. (2005). Meta-analysis: A unifying meta-likelihood approach framing unobserved heterogeneity, study covariates, publication bias, and study quality. *Methods of Information in Medicine, 44,* 127–136.

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

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Cheung, M. W. L. (2007a). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling, 14,* 227–246.

Cheung, M. W. L. (2007b). Comparison of methods of handling missing time-invariant covariates in latent growth models under the assumption of missing completely at random. *Organizational Research Methods, 10,* 609–634.

Cheung, M. W. L. (in press). Constructing approximate confidence intervals for parameters with structural equation models. *Structural Equation Modeling.*

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

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

Cheung, S. F. (2000). *Examining solutions to two practical issues in meta-analysis: Dependent correlations and missing data in correlation matrices.* Unpublished doctoral dissertation, Chinese University of Hong Kong.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10,* 101–129.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426–443.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Curran, P. J. (2002). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38,* 529–569.

Curran, P. J., & Bauer, D. J. (2007). Building path diagrams for multilevel models. *Psychological Methods, 12,* 283–297.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* New York: Cambridge University Press.

Demidenko, E. (2004). *Mixed models: Theory and applications.* Hoboken, NJ: Wiley-Interscience.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177–188.

Enders, C. K. (2001a). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61,* 713–740.

Enders, C. K. (2001b). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, 8,* 128–141.

Enders, C. K. (2006). A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine, 68,* 427–436.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods, 10,* 227–254.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8,* 206–213.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339–353.

Hafdahl, A. R. (2001). *Multivariate meta-analysis for exploratory factor analytic research.* Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Hall, D. B., & Wang, L. (2005). Two-component mixtures of generalized linear mixed effects models for cluster correlated data. *Statistical Modelling, 5,* 21–37.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hedges, L. V. (1982a). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7,* 119–137.

Hedges, L. V. (1982b). Fitting continuous models to effect size data. *Journal of Educational Statistics, 7,* 245–270.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93,* 388–395.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Russell Sage Foundation.

Hedges, L. V. (2007). Meta-analysis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 919–953). New York: Russell Sage Foundation.

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

Hedges, L. V., & Vevea, J. L. (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.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Erlbaum.

Hox, J. J., & de Leeuw, E. D. (2003). Multilevel models for meta-analysis. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 90–111). Mahwah, NJ: Erlbaum.

Huedo-Medina, T. B., Sánchez-Meca, F., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11,* 193–206.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8,* 275–292.

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

Jamshidian, M., & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24,* 21–41.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: A user's reference guide.* Chicago: Scientific Software International.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1,* 227–235.

Konstantopoulos, S. (2008). An introduction to meta-analysis. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 177–194). Thousand Oaks, CA: Sage.

Konstantopoulos, S., & Hedges, L. V. (2004). Meta-analysis. In M. Kaplan (Ed.), *Handbook of quantitative methods in the social sciences* (pp. 281–297). Thousand Oaks, CA: Sage.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). Boston: McGraw-Hill Irwin.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Little, R. J. A. (1992). Regression with missing *X*'s: A review. *Journal of the American Statistical Association, 87,* 1227–1237.

Lo, Y., Mendell, N., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10,* 21–39.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model, size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14,* 26–47.

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research, 41,* 499–532.

MacIntosh, R., & Hashim, S. (2003). Converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement, 27,* 372–379.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equation modeling. *Psychological Methods, 10,* 259–284.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Muthén, B. (2004a). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B. (2004b). Mplus technical appendices. Retrieved from http://www.statmodel.com/download/techappen.pdf

Muthén, B., & Asparouhov, T. (2002). Modeling of heteroscedastic measurement errors. Retrieved from http://www.statmodel.com/download/webnotes/mc3.pdf

Muthén, B., & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus. Retrieved from http://www.statmodel.com/download/webnotes/webnote6.pdf

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

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academy Press.

Neale, M. C. (2000). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples* (pp. 249–267). Hillsdale, NJ: Erlbaum.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). (Available from Virginia Commonwealth University, Department of Psychiatry, Box 900126, Richmond, VA 23298.)

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide.* Malden, MA: Blackwell.

Pigott, T. D. (1994). Methods for handling missing data in research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 163–175). New York: Russell Sage Foundation.

Pigott, T. D. (2001). Missing predictors in models of effect size. *Evaluation & the Health Professions, 24,* 277–307.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* Available from http://www.r-project.org

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Russell Sage Foundation.

Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin, 103,* 111–120.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Chichester, United Kingdom: Wiley.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Rovine, M. J., & Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 65–96). Washington, DC: American Psychological Association.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33,* 545–571.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Shadish, W. R. (1996). Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods, and issues. *Psychological Methods, 1,* 47–65.

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

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* Boca Raton, FL: Chapman & Hall.

Stapleton, L. M., & Leite, W. L. (2005). A review of syllabi for a sample of structural equation modeling courses. *Structural Equation Modeling, 12,* 642–664.

Stoel, R. D., Galindo-Garre, F., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods, 11,* 439–455.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research.* Chichester, United Kingdom: Wiley.

Sutton, A. J., & Pigott, T. D. (2005). Bias in meta-analysis: Induced by incompletely reported studies. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 223–239). Chichester, United Kingdom: Wiley.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52,* 393–408.

Takkouche, B., Cadarso-Suarez, C., & Spiegelman, D. (1999). Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *American Journal of Epidemiology, 150,* 206–215.

Thomas, H. (1989). A mixture model for distributions of correlation coefficients. *Psychometrika, 54,* 523–530.

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21,* 1559–1573.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18,* 2693–2708.

Van den Noortgate, W., & Onghena, P. (2005). Parametric and nonparametric bootstrap methods for meta-analysis. *Behavior Research Methods, 37,* 11–22.

Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91,* 217–221.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30,* 261–293.

Viechtbauer, W. (2006). MiMa: An S-Plus/R function to fit meta-analytic mixed-, random-, and fixed-effects models [Computer software and manual]. Retrieved from http://www.wvbauer.com/

Viechtbauer, W. (2007a). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine, 26,* 37–52.

Viechtbauer, W. (2007b). Hypothesis tests for population heterogeneity in meta-analysis. *British Journal of Mathematical and Statistical Psychology, 60,* 29–60.

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

Wang, X., Schumitzky, A., & D'Argenio, D. Z. (2007). Nonlinear random effects mixture models: Maximum likelihood estimation via the EM algorithm. *Computational Statistics & Data Analysis*, *51*, 6614–6623.

Xia, Y., Weng, S., Zhang, C., & Li, S. (2005). Mixture random effect model based meta-analysis for medical data mining. In P. Perner & A. Imiya (Eds.), *Machine learning and data mining in pattern recognition* (pp. 630–640). Berlin: Springer.

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: Elsevier/North-Holland.

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62,* 297–330.

(*Appendix follows*)

Appendix

Mplus Version 5 Codes for the Meta-Analysis

```
TITLE: Fixed-effects model: An intercept model
DATA: FILE IS hox.txt;
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter;
DEFINE: w2 = SQRT(varofd**(-1)); ! Weight for the transformation
      d = w2*d;                  ! Transformed d
      inter = w2*inter;          ! Transformed intercept
MODEL:
      d ON inter;
      [d@0.0];                   ! Intercept is fixed at 0
      d@1.0;                     ! Error variance is fixed at 1
OUTPUT: SAMPSTAT;

TITLE: Fixed-effects model: A continuous covariate
DATA: FILE IS hox.txt;
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter weeks;
DEFINE: w2 = SQRT(varofd**(-1));
      d = w2*d;
      inter = w2*inter;
      weeks = w2*weeks;          ! Transformed weeks
MODEL:
      d ON inter weeks;
      [d@0.0];                   ! Intercept is fixed at 0
      d@1.0;                     ! Error variance is fixed at 1
OUTPUT: SAMPSTAT;

TITLE: Random-effects model (with ML method): An intercept model
DATA: FILE IS hox.txt;
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter;
DEFINE: w2 = SQRT(varofd**(-1));
      d = w2*d;
      inter = w2*inter;
ANALYSIS: TYPE=RANDOM;           ! Use random slope analysis
MODEL:
      [d@0.0];                   ! Intercept is fixed at 0
      d@1.0;                     ! Error variance is fixed at 1
      u | d ON inter;            ! u: random effects
      u*;                        ! var(u): tau^2
      [u*];                      ! mean(u): weighted effect size
OUTPUT: SAMPSTAT;

TITLE: Random-effects model (with ML method): A continuous covariate
DATA: FILE IS hox.txt;
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter weeks;
DEFINE: w2 = SQRT(varofd**(-1));
      d = w2*d;
      inter = w2*inter;
      weeks = w2*weeks;
```

```
ANALYSIS: TYPE=RANDOM;            ! Use random slope analysis
MODEL:
      [d@0.0];                    ! Intercept is fixed at 0
      d@1.0;                      ! Error variance is fixed at 1
      u | d ON inter;
      d ON weeks;
      u*;                         ! var(u): tau^2
      [u*];                       ! mean(u): intercept
OUTPUT: SAMPSTAT;
```

TITLE: Fixed-effects model: Missing data on the covariate (FIML)

```
DATA: FILE IS hox_miss.txt;       ! Data file with missing values
      ! FIML is the default option started from Mplus Version 5
      ! LISTWISE = ON;            ! Use listwise deletion
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter weeks;
      MISSING ARE ALL (999);      ! Define missing values
DEFINE: w2 = SQRT(varofd**(-1));
      d = w2*d;
      inter = w2*inter;
      weeks = w2*weeks;           ! Transformed weeks
MODEL:
      d ON inter weeks;
      [d@0.0];                    ! Intercept is fixed at 0
      d@1.0;                      ! Error variance is fixed at 1
OUTPUT: SAMPSTAT;
```

TITLE: Fixed-effects model: An intercept model with Q statistic and heterogeneity indices

```
DATA: FILE IS hox.txt;
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter;
DEFINE: w2 = SQRT(varofd**(-1));
      d = w2*d;
      inter = w2*inter;
ANALYSIS: BOOTSTRAP=2000;         ! Use bootstrap analysis
MODEL:
      d ON inter;
      [d@0.0];                    ! Intercept is fixed at 0
      d (p1);                     ! Estimated error variance
MODEL CONSTRAINT:
      NEW(Q_stat H2_stat I2_stat);
      Q_stat = 20*p1;             ! Q statistic
      H2_stat = Q_stat/19;        ! H2 index
      I2_stat = 1-19/Q_stat;      ! I2 indice
OUTPUT: SAMPSTAT;
      CINTERVAL(BCBOOTSTRAP);     ! Bias-corrected bootstrap CI
      ! CINTERVAL(symmetric);     ! Wald CI
```

TITLE: Fixed-effects model: An intercept model with two mixtures

```
DATA: FILE IS hox.txt;
VARIABLE: NAMES d varofd inter weeks;
      USEVARIABLES ARE d inter;
      CLASSES=c(2);               ! Define two classes of mixtures
DEFINE: w2 = SQRT(varofd**(-1));
```

(*Appendix continues*)

```
      d = w2*d;
      inter = w2*inter;
ANALYSIS: TYPE=MIXTURE;           ! Use mixture analysis
      STARTS 200 20;              ! Use 200 random starting values
      STITERATIONS = 20;
MODEL:
      %OVERALL%                   ! Overall model
      [d@0.0];                    ! Intercept is fixed at 0
      d@1.0;                      ! Error variance is fixed at 1

      d ON inter;                 ! Weighted mean of the 1st mixture

      %c#2%                       ! Second mixture
      d ON inter;                 ! Weighted mean of the 2nd mixture
OUTPUT: SAMPSTAT;
      TECH11;                     ! Request an aLRT
SAVEDATA:
      SAVE=CPROB;                 ! Save the posterior probabilities
      FILE IS intercept_2mix.txt;
```

Received June 5, 2007
Revision received May 16, 2008
Accepted May 29, 2008 ■