5-1-2010

An Evaluation of Multiple Imputation for Meta-Analytic Structural Equation Modeling

Carolyn F. Furlow
Georgia State University, cfurlow@cdc.gov

S. Natasha Beretvas
The University of Texas at Austin, tasha.beretvas@mail.utexas.edu

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

 Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

An Evaluation of Multiple Imputation for Meta-Analytic Structural Equation Modeling

Carolyn F. Furlow
Georgia State University

S. Natasha Beretvas
University of Texas at Austin

A simulation study was used to evaluate multiple imputation (MI) to handle MCAR correlations in the first step of meta-analytic structural equation modeling: the synthesis of the correlation matrix and the test of homogeneity. No substantial parameter bias resulted from using MI. Although some SE bias was found for meta-analyses involving smaller numbers of studies, the homogeneity test was never rejected when using MI.

Key words: Meta-analysis, structural equation modeling, multiple imputation, missing data.

Introduction

Meta-analytic structural equation modeling (MASEM) has been recommended as a useful approach for supporting theoretical models and combines the benefits of both meta-analysis and structural equation modeling (SEM). The meta-analytic benefits include the use of quantitative synthesis methods which allow a researcher to cull correlations from multiple studies that can then be combined across those studies to provide individual, more precise estimates of each relevant correlation. This can be conducted for each element of a correlation matrix that describes the full set of relationships between the variables of interest to the MASEM researcher. The resulting meta-analytically pooled correlation matrix can then be analyzed using SEM procedures.

Several methodological dilemmas and impediments are frequently encountered by MASEM researchers. Most commonly, applied

Carolyn F. Furlow conducted this study while in the Department of Educational Policy Studies. Current email: cfurlow@cdc.gov. Natasha Beretvas is an Associate Professor and chair of Quantitative Methods in the Department of Educational Psychology. Her interests are in multilevel and meta-analytic modeling techniques. Email her at: tasha.beretvas@mail.utexas.edu.

researchers synthesize elements of a correlation matrix which is then analyzed using SEM software. The MASEM researcher is likely to encounter problems with missing data. In the MASEM context, this can be in the form of missing studies or missing correlations (Pigott, 1994). The focus of this study concerns performance of multiple imputation for handling missing correlations for the first step of MASEM, the synthesis of the correlation matrices across studies.

Missing Data in MASEM
    If a researcher were interested in summarizing elements of a correlation matrix describing relationships among five variables, ideally data from each contributing study would include estimates of each of the correlations in the matrix. This is rarely the case. At the primary study level, several possible reasons exist to explain why a correlation might not be reported. The authors of the study might not have been interested in measuring one of the five variables of interest to the meta-analyst, or at the time when one of the primary studies was conducted, a variable of interest to the meta-analyst might not yet have been conceptualized as a construct that exhibits an interesting relationship with other variables in the matrix (Furlow & Beretvas, 2005). Thus, in either scenario, the study would not include correlations of that variable with each of the remaining four.

Another plausible reason why a correlation might be missing from a study may be the file drawer problem (Rosenthal, 1979), most commonly referred to as publication bias. Publication bias describes the tendency of authors (and editors) to provide statistical results (either descriptive or inferential) only for statistically significant results. Authors either fail to mention uninteresting (commonly meaning statistically non-significant) results, or journal space limitations restrict the presentation of the relevant values offering only the phrase "not statistically significant".

Authors using MASEM reported using a variety of methods for handling missing correlation estimates. Hom, et al. (1992) utilized listwise deletion (LD) by only incorporating results from studies that provided the full set of correlations of interest. The vast majority of MASEM researchers (e.g., Brown & Stayman, 1992; Conway, 1999; Manfredo, et al., 1996; Parker, et al., 2003; Premack & Hunter, 1988; Tett & Meyer, 1993) used pairwise deletion (PD). A few used single value imputation to handle missing correlations in their MASEM studies (Bailey, 2001; Colquitt, LePine & Noe, 2000). The single imputation method involves either mean imputation (using the mean of the correlation estimates provided in other studies in the meta-analysis) or substituting a value based on related results from other meta-analytic research conducted outside the domain of the focal MASEM study (Colquitt, et al., 2000). It is unclear in some MASEM articles how the missing correlations were handled (e.g., Verhaeghen & Salthouse, 1997). To date, no applied study has used multiple imputation (MI) to handle missing correlations.

As with any statistical analysis, the source of the missing data impacts how well the method used to handle the missing data will function. Little and Rubin (1987) categorized missing data mechanisms into three types: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). What distinguishes these missing data mechanisms is the relationship between the missing (unreported or unobserved) data point and the complete set of data. In traditional statistical analyses the cases (rows) in a dataset are a single study's participants and the data points are participants' scores on each variable (columns). In meta-analysis, the columns (variables) consist of correlation estimates for each row, which represents each study.

With MCAR data, the correlation's missingness is unrelated to any of the observed correlations in the dataset. As an example, when data are not gathered in a primary study due to lack of funding and that funding is not related to any of the variables and thus to correlations in the dataset, then the missingness can be considered MCAR (Pigott, 1994). Another example of MCAR data occurs when a primary researcher does not measure a variable of interest for the MASEM because it is not theoretically relevant to his/her study and thus that variable's correlations with the other variables would be missing.

With MAR data, a correlation's missingness ($M_{r_{ij}}$ = 1 if correlation r_{ij} is missing, and $M_{r_{ij}}$ = 0, otherwise) would be related to the correlation's value but only indirectly - specifically only through another observed correlation (or correlations) in the dataset (Collins, Schafer & Kam, 2001). For example, the likelihood of a correlation, r_{WX}, being missing in a study might be greater for higher values of another observed correlation, r_{YZ}. There would then appear to be a simple relationship between r_{WX} and $M_{r_{WX}}$. However, if within levels of r_{YZ}, values of r_{WX} are unrelated to $M_{r_{WX}}$, then the data are MAR. In other words, if the relationship between r_{WX} and its likelihood of being missing (i.e., $M_{r_{WX}}$) is fully explained by the relationship between r_{YZ} and $M_{r_{WX}}$, then the missingness can be considered MAR.

As a meta-analytic example, studies being synthesized might involve an assessment of the relationship between constructs W, X, Y and Z. There might be a variety of scales that are designed to assess each of Y and Z. Researchers who use certain (more reliable psychometrically) measures of Y and Z might espouse a theoretical framework that also means they are more likely to be interested in the relationship between measures X and W. Researchers who use

130

different measures of *Y* and *Z* may be less likely to assess *X* and *W*. When using more reliable measures of *Y* and *Z*, the resulting r_{YZ}s will tend to be stronger than the r_{YZ}s based on less reliable scores. Thus, for higher values of r_{YZ}, the likelihood that r_{XW} is reported is higher than for lower values of r_{YZ}. And for lower values of r_{YZ}, it is more likely that r_{XW} will be missing. However, controlling for r_{YZ}, there is no relationship between r_{XW} and the likelihood that r_{XW}'s value is missing from a study. Thus missing r_{XW} values could be considered MAR.

MNAR data result when the likelihood of a missing correlation is related to the value of the (missing) correlation itself. Publication bias provides a likely cause of MNAR data. As mentioned earlier, if a correlation estimate is not statistically significant, an author might not report the relevant statistical information and/or an editor might censor the presentation of such results. If publication bias explains the missingness, then the likelihood of missingness is negatively related to the correlation estimate's value, all other factors being equal. The opposite pattern of MNAR (in which there is a positive relationship between the missing correlation's value and the likelihood of its being missing) is also possible. It can occur when a researcher purposely neglects to report a correlation that is stronger than would be expected theoretically.

Use of listwise deletion (LD) to handle missing data can be advocated in situations in which only a few data points are missing. LD has been found to result in unbiased parameter estimates for models estimated with MCAR data (Allison, 2003). However, LD can also result in a drastic reduction in statistical power under conditions with high proportions of missing data. Graham and Hofer (2000) recommend that if only five percent or less of the dataset is MCAR, then LD can be used. Unfortunately, LD is usually not a feasible alternative in MASEM research. A high proportion of study authors do not report all correlations of interest to MASEM researchers (Furlow & Beretvas, 2005), for example, in Premack and Hunter's (1988) MASEM study, if LD had been used it would have resulted in a completely empty dataset.

Many MASEM researchers do not use LD but instead employ available case analysis (PD) as the preferred method for handling missing correlations (Furlow & Beretvas, 2005). When using PD, no information is deleted; each element of the correlation matrix is instead obtained by synthesizing all available, observed correlation estimates. Use of PD with conventional (i.e., not meta-analytic) data has been found to result in approximately unbiased parameter estimates for MCAR data, however, PD can lead to biased estimates if data are MAR or MNAR (Graham & Hofer, 2000). Use of PD has also been found to lead to non-positive definite correlation matrices for typical, non-meta-analytic datasets (Arbuckle, 1996; Graham & Hofer, 2000). To date, this problem has been reported in only one applied MASEM study (Kubeck, 2002). Even the few MASEM simulation studies that have been conducted to evaluate the performance of PD with missing data have not encountered non-positive definite matrices (S. F. Cheung, 2000; M. Cheung & Chan, 2005; Furlow & Beretvas, 2005).

An additional problem associated with PD is encountered when PD is used to calculate a correlation matrix for a conventional SEM analysis (Allison, 2003) and when PD is used to calculate elements of a synthesized correlation matrix to be analyzed using MASEM. In the SEM scenario, each element of the correlation matrix might be based on different sample sizes and yet a single sample size must be associated with the matrix used to estimate the structural equation model. The same dilemma is encountered by MASEM researchers who use the synthesized correlation matrix in their SEM analysis (without the associated covariance matrix for the correlations). In Cheung and Chan's MASEM procedure utilizing the covariance matrix, the authors assert that use of the total sample size is "free from the ambiguity of choosing among different sample size values that have been proposed" (2005, p. 47); however, it is unclear that this is the case.

Another method to handle missing data in MASEM research could be through the use of mean imputation to impute a missing data point's value (Graham & Hofer, 2000). The problem with mean imputation is that it deflates the associated variability of the relevant estimate (the correlation in MASEM); this holds even when the missing data mechanism is MCAR, thus mean imputation is not recommended. To

compensate for the reduction in variability, it is possible to use Bayesian multiple imputation (MI) procedures (Rubin, 1978; Rubin, 1987; Schafer, 1997). MI has not been used in meta-analysis in general and the goal of the current study is to investigate its use with MASEM.

Multiple Imputation

No applied MASEM study to date has examined the performance of multiple imputation (MI); however, MI is a promising technique for handling missing data found in MASEM research. MI expands upon single imputation and its resultant attenuation of variability. MI takes into account the uncertainty involved in missing data and imputes m plausible values (where $m > 1$) to replace each single missing data point (each correlation estimate in MASEM research) resulting in m datasets. Each imputed dataset will have the same values for the non-missing correlation estimates. The values imputed for the missing data points will distinguish the m datasets. Each of the m datasets is analyzed using the statistical procedure of interest (i.e., the meta-analysis) and the results can be summarized across the imputed datasets. To obtain unbiased estimates using MI, the missing data are assumed to be at most MAR (thus, MI will also work well with MCAR data).

MI uses the Bayesian Markov chain Monte Carlo (MCMC) algorithm to impute values for missing data points. The reader is referred to several excellent chapters, texts and articles that provide more information on the technical process underlying MI (Allison, 2003; Peugh & Enders, 2004; Graham & Hofer, 2000; Schafer & Graham, 2002; Schafer & Olsen, 1998).

In traditional use of MI, the researcher calculates the statistic of interest (whether it is a sample mean, a correlation, a regression coefficient, etc.) represented generally as $\hat{q}_i$ for imputed dataset i. The simple average of the m estimates can be combined across imputed datasets to provide the multiply imputed estimate of the statistic using:

$$\bar{q} = \frac{1}{m}\sum_{i=1}^{m}\hat{q}_i \qquad (1)$$

(Rubin, 1987). The variance estimate associated with $\bar{q}$ is a function of the variance within each imputed dataset

$$\bar{u} = \frac{1}{m}\sum_{i=1}^{m}u_i \, , \qquad (2)$$

as well as the variability between the imputed datasets (Rubin, 1987)

$$b = \frac{1}{(m-1)}\sum_{i=1}^{m}[\hat{q}_i - \bar{q}]^2 \, . \qquad (3)$$

The total variance can be calculated using

$$t = \bar{u} + \frac{1}{(1+m)}b \, . \qquad (4)$$

In MASEM, the meta-analysis involves synthesizing correlations across studies. If MI were used, values for correlations would be imputed leading to the construction of m complete (imputed) datasets of correlations. A synthesized correlation is calculated for each correlation (e.g., r_{WX}, r_{WZ}, r_{XZ}, etc.) in each dataset and each resulting synthesized correlation corresponds to the relevant $\hat{q}$ (previously mentioned), thus, equations 1 - 4 can be used to calculate the MI estimate of each synthesized correlation and its associated variance.

Although parameter and standard error estimates can be easily combined using Equations 1 - 4, multivariate inferences, such as the test of homogeneity in meta-analysis, require different formulas. For example, Schafer's (1997) formula for combining χ^2 values (such as the one from the test of homogeneity) across studies is a relatively simple function of each imputation's χ^2 statistic value and its df. The formula provides an F-ratio statistic for which an associated p-value can be estimated that can be interpreted as the significance test associated with the χ^2. The formula is:

$$F(df_{\chi^2}, df_{Error}) = \frac{\left[\dfrac{\overline{\chi}^2}{\left(df_{\chi^2}\right)}\right] - \left[r_2\dfrac{(m+1)}{(m-1)}\right]}{1+r_2} \quad (5)$$

where df_{χ^2} is the df associated with the χ^2, $\overline{\chi}^2$ is the mean of the m imputations' χ^2 values, df_{Error} is the error degrees of freedom of the F-ratio statistic calculated as follows:

$$df_{Error} = \frac{(m-1)\left(1+r_2^{-1}\right)^2}{\left(df_{\chi^2}\right)^{3/m}} \quad (6)$$

and (Equation 5), r_2 is the sample variance of $\sqrt{\chi^2}$ across imputations where r_2 is calculated as follows:

$$r_2 = \frac{(1+1/m)}{(m-1)}\left[\sum_{i=1}^{m}\chi_i^2 - \frac{\left(\sum_{i}^{m}\sqrt{\chi_i^2}\right)^2}{m}\right]. \quad (7)$$

Rubin also derived the formula for the efficiency of estimates based on m imputed datasets

$$(1+\gamma/m)^{-1} \quad (8)$$

where γ is the fraction of missing information. In most cases between five and ten imputations are sufficient to achieve efficient results, however, with a large degree of missingness, more imputations may be necessary (Allison, 2003; Hershberger & Fisher, 2003). With a higher number of imputations, estimates of parameters become more stable (Allison, 2003). Allison notes that one diagnostic test of whether more imputations are necessary requires a check of the degrees of freedom for each parameter estimate. If the degrees of freedom are appreciably below 100 then more imputations should improve the efficiency of the estimates.

Typical MI procedures assume that data are multivariately normal. In MASEM, the typical unit of analysis is the correlation, r. The sampling distribution of rs sampled from non-zero ρ, however, tends to be increasingly skewed for larger $|\rho|$ (Hedges & Olkin, 1985). The use of Fisher's Z_r transformation results in a more normal sampling distribution even for larger ρ (Steiger, 1980). While the resulting Z_rs are approximately normally distributed, research has suggested that MI is reasonably robust to violations of the assumption of normality (Enders, 2001; Graham & Schafer, 1999). Graham and Schafer's (1999) simulation study found that - even for extremely non-normal variables and small sample sizes - MI worked very well.

A benefit of using MI to handle missing data involves the less restrictive MAR missingness mechanism that can be assumed (unlike with PD and LD where only MCAR is assumed). Maximum likelihood (ML) methods and the expectation maximization (EM) algorithm, among others, also offer alternatives for handling missing data (Collins, Schafer & Kam, 2001). Use of MI, however, is less computationally intensive than ML (Sinharay , et al., 2001) and most MI programs use the EM algorithm to estimate starting values for the ensuing data augmentation iterations. Use of MI is further facilitated by its availability in several software packages including NORM (Schafer, 1999), SAS PROC MI (SAS Institute, 2005) S-PLUS (version 6.0, Insightful Corporation, 2001), and SPSS (version 14.0, SPSS, 2006).

To date, no meta-analytic researchers have used MI when handling missing data. The focus of the current study is to evaluate use of MI for synthesizing correlation matrix elements and their corresponding standard errors for use in MASEM. After missing correlations have been handled in MASEM, the researcher can synthesize the correlation matrix elements across studies. Before this synthesizing can occur, however, the researcher must decide whether to synthesize the correlations univariately or multivariately.

Synthesizing Correlations: Univariate Synthesis

Several methods exist that are used to synthesize effect sizes (here, correlations) across the k studies included in a meta-analysis (Hedges & Olkin, 1985). Synthesis methods typically involve weighting each effect, e, by a function of its associated sample size. The weight, w, most commonly used to obtain the pooled estimate of the effect size, $\hat{\varepsilon}$, is the inverse of the effect's conditional variance (Cooper & Hedges, 1994):

$$\hat{\varepsilon} = \frac{\sum_{i=1}^{k} w_i e_i}{\sum_{i=1}^{k} w_i} . \qquad (9)$$

This weighting assigns more weight to the more precise correlation estimates that are associated with larger sample sizes.

As noted, the sampling distribution of r is increasingly skewed for larger values of $|\rho|$ (Hedges & Olkin, 1985). In addition, the large-sample variance of this distribution depends on the value of the parameter itself (Becker, 2000). For this reason, several meta-analytic researchers and in particular MASEM researchers (for example, Becker & Fahrbach, 1994; Hafdahl, 2001; 2007) recommend using Fisher's (1928) r-to-Z_r normalizing and variance-stabilizing transformation:

$$Z_r = .5\{\ln[(1+r_i)/(1-r_i)]\} \qquad (10)$$

when synthesizing correlation estimates. The variance estimate associated with Z_r is:

$$\hat{\sigma}_{Z_r}^2 = \frac{1}{n_i - 3} . \qquad (11)$$

The weight associated with the correlation estimate for study i is thus $w_i = (n_i - 3)$, thus, to obtain the pooled estimate of the transformed correlation, $\hat{\zeta}_\rho$, between variables X and Y, the following equation is used:

$$\hat{\zeta}_\rho = \frac{\sum_{i=1}^{k} (n_i - 3) Z_{r_i}}{\sum_{i=1}^{k} (n_i - 3)} \qquad (12)$$

The resulting pooled, transformed correlation estimate, $\hat{\zeta}_\rho$, is then back-transformed to the correlation metric using the back-transformation formula:

$$\hat{\rho} = \frac{\exp(2\hat{\zeta}_\rho) - 1}{\exp(2\hat{\zeta}_\rho) + 1} \qquad (13)$$

to obtain the pooled estimate of the correlation, $\hat{\rho}$. This univariate synthesis method can be used for each correlation in the matrix of interest. The standard error of $\hat{\zeta}_\rho$ is calculated using

$$s_{\hat{\zeta}_\rho} = \sqrt{\frac{1}{\sum_{i=1}^{k} n_i - 3}} . \qquad (14)$$

Synthesizing Correlations: Multivariate Synthesis

Becker (1992b) introduced a multivariate synthesis method using generalized least squares (GLS) estimation that recognizes the possible dependencies among the $p*$ effect sizes (where, $p* = [p(p-1)]/2$ correlations among p variables in a correlation matrix):

$$\hat{\theta} = \left(X'\Sigma^{-1} X \right)^{-1} X'\Sigma^{-1} T \qquad (15)$$

where $\hat{\theta}$ is a $p*$ x 1 column vector containing the multivariately synthesized estimates of the $p*$ effect sizes, X is the design matrix consisting of k stacked $p*$ x $p*$ that identifies matrices for $p*$ effect sizes per study, Σ is a block-diagonal matrix containing the covariance matrix for each study's set of effect sizes as blocks along its diagonal and T is a $kp*$ x 1 column vector containing each study's effect size estimates. The omnibus Q-statistic is used to test the null hypothesis of the homogeneity of effect sizes

(correlation matrices) across studies. It can be calculated using:

$$Q = T'\Sigma^{-1}T - \hat{\theta}'\left(X'\Sigma^{-1}X\right)\hat{\theta} \qquad (16)$$

and is assumed to follow a χ^2 distribution with $(k-1)p^*$ degrees of freedom.

Olkin and Siotani's (1976) large-sample approximation to the covariance between two correlations should be used to calculate elements of Σ in equations 15 and 16:

$$\sigma_{r_{ist},r_{iuv}} = [0.5\rho_{ist}\rho_{iuv}(\rho_{isu}^2 + \rho_{isv}^2 + \rho_{itu}^2 + \rho_{itv}^2)$$
$$+ \rho_{isu}\rho_{itv} + \rho_{isv}\rho_{itu} - (\rho_{ist}\rho_{isu}\rho_{isv}$$
$$+ \rho_{its}\rho_{itu}\rho_{itv} + \rho_{ius}\rho_{iut}\rho_{iuv}$$
$$+ \rho_{ivs}\rho_{ivt}\rho_{ivu})]/n_i$$

$$(17)$$

Alternatively, multivariate synthesis with GLS estimation (see equation 15) could be used to synthesize Z_r-transformed correlations (equation 10). Elements of the covariance matrix, Σ, for the Z_rs are a function of the covariances between the correlations (equation 17) and can be calculated using:

$$\sigma_{Z_{r_{ist}},Z_{r_{iuv}}} = \frac{n_i\sigma_{r_{ist},r_{iuv}}}{(n_i - 3)\left(1 - \rho_{ist}^2\right)\left(1 - \rho_{iuv}^2\right)} \qquad (18)$$

(Steiger, 1980). Initially, when demonstrating use of GLS synthesis in a simulation study (Becker, 1992b), individual study estimates of ρ were used in equations 17 or 18 when calculating Σ. Use of these less efficient single study estimates of ρ was later found to be one cause of GLS' poor performance for synthesizing correlation matrix elements (Becker & Fahrbach, 1994).

Researchers have found that multivariate GLS tends to outperform univariate synthesis methods when a pooled estimate of ρ is instead substituted for each ρ in equations 17 or 18 (Becker & Fahrbach, 1994; S. Cheung, 2000; Furlow & Beretvas, 2005). Lastly, it should be mentioned that when there is no missingness, or the missing values have been replaced using some type of imputation, the results with Z_r-transformed correlations from GLS and univariate synthesis are mathematically equivalent (Gagné, Furlow, & Beretvas, 2004).

Hafdahl (2007) conducted a study evaluating the performance of univariate and multivariate synthesis methods paired with r and Fisher's transformation, Z_r, as well as using the more efficient estimates of ρ in the relevant weight (univariate or multivariate) matrix. Hafdahl found support for using the Z_r transformation over r, for multivariate over univariate synthesis methods and for substituting the pooled estimates of ρ instead of using individual study estimates. Combining these options led to better parameter estimation accuracy, efficiency and precision and for improved Type I error control for the test of homogeneity.

Hafdahl (2007) only investigated the performance of synthesis methods when no data (i.e., correlation estimates) were missing in any of the studies being meta-analyzed. In cases where not all correlation estimates are provided in every study, the relevant rows and columns are deleted from the matrices (specifically in the T, Σ, and X matrices) used in GLS (equation 15) and the Q-statistic (equation 16) estimation (Becker & Schram, 1994). Other researchers have assessed the impact of missing data on MASEM. Similar to Hafdahl's results, Furlow and Beretvas (2005) found that the Z_r transformation along with use of pooled average estimates of ρ substituted for ρ in the elements of the Σ matrix worked best as a synthesis method. Furlow and Beretvas (2005) also compared the results from their study when correlations were MCAR and MNAR and when LD versus PD was used to handle the missing correlations. They found that MNAR data produced high levels of relative bias in the correlation estimates while percent relative bias among the correlations for MCAR data was never above 3%. Use of PD resulted in enhanced estimation of synthesized correlations when compared with LD when it was used along with the more efficient method for GLS.

Cheung and Chan (2005) demonstrated the use of multi-group SEM (where each study comprises a group) and model parameter constraints across groups as a way to conduct

MASEM analyses. They compared the performance of their procedure with the more typically used MASEM procedure (in which the elements in the correlation matrix are first meta-analytically synthesized and then analyzed using canned SEM software without the covariance matrix associated with the synthesized correlation matrix being analyzed).

The authors evaluated both procedures when data were missing and found support for their procedure, however, the authors used the earlier version of GLS with individual study estimates of ρ when calculating the covariance matrix, Σ, used in GLS' multivariate synthesis (equation 15) and for the Q-statistic (equation 16). Therefore, it was not surprising that GLS did not perform well. In addition, Cheung and Chan only considered $k = 5$, 10 and 15 in their study (well below the median k of 26 that they reported in their review of the applied literature). Although they acknowledged that their largest k was smaller than the average reported, they indicated that because their method involved the cross-group constraints (where each study is considered as a group) their method was too computationally intensive to involve larger ks, thus providing a weakness to their method for MASEM.

Both methodological MASEM studies (Furlow & Beretvas, 2005; Cheung & Chan, 2005) had also generated data such that the sample size associated with each study was the same. While use of a single sample size for every study in the simulated meta-analysis might simplify interpretation of results, it does not provide an authentic simulation of reality. Instead, in a real-world meta-analysis the sample size for each study is typically different.

Methodology

This simulation study was designed to investigate the use of MI for pooling estimates of correlation matrices when some correlation estimates were missing in the primary studies being synthesized. For this exploration of the use of MI, the synthesis of elements of a four-variable correlation matrix was investigated with MCAR data. Manipulated conditions in the study included the degree of missingness (25% and 50% of all correlations), the number of studies ($k = 25$, 50 and 100) involved in the

meta-analysis, and the average sample size per study ($\tilde{n} = 50$ and 100).

Because MI assumes data are multivariate normal, it was of interest in this study to transform the correlations to Fisher's Z metric since its sampling distribution is more normal than that of ρ. As noted, when Z_r is used, results from the more efficient version of GLS are equivalent to using univariate weighting when no data are missing or missing data values have been imputed (Gagné, Furlow & Beretvas, 2004). The results from the omnibus test of the homogeneity of variance, however, will not be the same for univariate and multivariate synthesis. Thus, the performance of univariate versus multivariate synthesis methods was compared when assessing the Type I error performance for the homogeneity test. Use of MI was also assessed in terms of resulting parameter and standard error estimation for only the univariate synthesis of correlations.

In applied meta-analysis, study results tend to be based on uneven sample sizes. To mimic this, each study's sample size, n_i, was generated using the same distribution as that used in Hafdahl's (2007) simulation study:

$$ n_i = \left\{ \left(\frac{\tilde{n}}{2} \right) \left[\frac{(X_i - 3)}{\sqrt{6}} \right] + \tilde{n} \right\} \qquad (19) $$

where X_i (for $i = 1$ to k) was sampled for each study i from a χ^2 distribution with 3 degrees of freedom. The value of $\tilde{n}$ depended on the sample size condition. In the current study, for small and moderate $\tilde{n}$ conditions, the values of $\tilde{n}$ were 50 and 100, respectively. Last, $\{y\}$ represents the closest integer to the value of y.

Data Generation

Multivariate normal fixed-effects data were generated in SAS (SAS Institute, 2005) using the Cholesky root of the generating covariance matrix. For each combination of conditions, 1,000 replications were conducted. The relevant degree of missingness was introduced into the dataset, the missingness was then handled using MI, and correlation estimates were synthesized.

Model Generation

To simplify this exploration of the use of MI with MASEM analyses, a scale-invariant model was selected to generate the data. Scale-free parameter and standard error estimation results for a scale-invariant SEM model are equivalent whether a correlation or a covariance matrix is analyzed (Cudeck, 1989). A four-variable, one-factor (scaled to have a variance of one) model was used. Values for the elements of the correlation matrix used to generate the data are those implied by the relevant generating values for the true factor loadings (with loading values of 0.5, 0.6, 0.7 and 0.8 for variables V1, V2, V3 and V4, respectively). Table 1 shows the model-implied values of the correlations used to generate the data. To simplify the study, the variables' and factors' variances were each standardized to have a value of one.

Table 1: Generating Values of Model-Implied Correlation Matrix

	V1	V2	V3	V4
V1	1.00			
V2	0.30	1.00		
V3	0.35	0.42	1.00	
V4	0.40	0.48	0.56	1.00

Note: Corresponding generating loading values were 0.5, 0.6, 0.7 and 0.8 for V1, V2, V3 and V4, respectively.

After data were generated, MCAR missingness was introduced. The premise underlying the general design of missingness in this simulation study was that once a study was (randomly) selected to have missingness introduced, then a variable was randomly selected as one that was not measured in a study. Once a variable was selected to be missing, all correlations involving that variable were designated as missing. Thus, if variable V1 was selected then r_{12}, r_{13}, r_{14}, would each be missing for a study: this mimics a realistic meta-analytic scenario in which a variable is not measured in a study and thus associated correlations could not be reported.

To determine reasonable values for the degree of missing data and the number of studies synthesized in the meta-analysis for this study, a review of applied MASEM studies in the literature was conducted through a search of the PsycInfo database using the search criteria "meta-analysis" and each of "structural equation modeling", "path analysis" and "confirmatory factor analysis". In addition, other applied MASEM articles cited in the resulting sources or known to the authors were also examined. This led to the identification of 24 applied MASEM studies. The amount of missing correlations could only be determined for 13 of these 24 studies because authors did not report the information needed to calculate these percentages. Two studies reported no missing correlations while at the other extreme, three studies reported over 80% of all correlations missing. The mean rate of missing correlations was 67.8% while the median rate was 70%. The mean number of studies synthesized across all 24 MASEM studies was 49.6 with correlations being pooled from a minimum of four to a maximum of 155 studies' results.

The number of studies in the simulated meta-analysis used in this study were chosen to reflect small (25), moderate (50) and large (100) numbers of studies. Per-study sample size was varied as described in equation 19 with two levels for the average per-study sample size ($\tilde{n}$ = 50 and 100). Two percentage levels of missing correlations were chosen (25% and 50%) to reflect the amounts of missingness found in applied MASEM studies. In conditions where 25% of the correlations were missing in a meta-analysis, 30% of the studies were first selected and then 50% (2) of the four variables within those studies were chosen to be missing (resulting in one correlation out of six remaining in those studies). In conditions with 50% of correlations missing, 60% of the studies were selected to have missingness and 50% of the variables within those studies were designated as missing along with their correlations. Baseline conditions where no correlations were missing were also examined for each combination of k and $\tilde{n}$.

Synthesis Method

Following dataset generation for each condition and introducing the missingness, MI was used. To multiply impute the data, SAS PROC MI was utilized, employing a noninformative prior (the default for PROC MI) and assuming a multivariate normal posterior distribution. Because a relatively large degree of missingness was simulated in this study, forty imputations were used rather than the typical five to ten (Allison, 2003). The forty imputed datasets each consisted of a full set of correlations for each study in each simulated meta-analysis.

After the forty multiply imputed datasets had been calculated for each replication the contents of each dataset were used to obtain forty synthesized matrices. To synthesize each correlation estimate, each study's r_{ij} value was transformed to $Z_{r_{ij}}$ using equation 10. These were pooled together using equation 12 to obtain the univariately pooled $\hat{\zeta}_{\rho_{ij}}$ for each pair of variables i and j. The standard error estimates were also calculated using equation 14. The resulting estimates of the Fisher-transformed correlation matrix elements and associated standard error estimates were then combined across the 40 imputed datasets per replication using Rubin's combination rules (see equations 1-4) through PROC MIANALYZE.

Performance of the Q-statistic for correct identification of the homogeneity of the correlation matrices across studies was also evaluated. The Q-statistic was calculated with the covariance matrix, Σ, in equation 16 containing only variances of the Z_rs along the diagonal for the test of homogeneity for the univariate synthesis. The Q-statistic was also calculated using the full covariance matrix (see equations 17 and 18) for GLS. Rather than using single-study estimates of ρ, the more efficient pooled estimates were used because they have been found to enhance the performance of GLS (Hafdahl, 2007). Per-imputation χ^2 estimates and associated p-values were combined across imputations using Allison's SAS macro COMBCHI (2000) (which utilizes equations 5, 6, and 7). A correction to COMBCHI corrected

a small error in the code (Enders, personal communication, December 8, 2005).

Data Analysis

The relative percent bias, $B(\hat{\theta})$, was used to evaluate estimation of correlations (Hoogland & Boomsma, 1998). Hoogland and Boomsma recommended identification of bias when the magnitude of $B(\hat{\theta})$ exceeds five percent of the corresponding population value. The accuracy of the standard error estimates associated with each correlation was assessed using the standard error's relative percent bias. Hoogland and Boomsma suggested that standard error relative percent bias of magnitude 10 percent or less indicates an acceptable degree of bias. Finally, the proportion of correct fixed-effects model identifications were tallied using the univariately and multivariately weighted Q-statistic (see equation 16).

Results

Parameter Estimation Bias

No substantial bias was found under any of the conditions examined for estimation of the correlations. Relative percent bias for each element across conditions and matrices never exceeded a magnitude of 1%.

Standard Error Estimation Bias

Table 2 lists the results from all study conditions and all correlations for the standard error bias. In cells with no missing correlations, percent relative bias was always well below Hoogland and Boomsma's (1998) 10% cutoff with a highest magnitude of 5.5%. In cells with missing data, the bias was always positive and a distinction was apparent in the bias for the small (25%) and large (50%) degree of missingness conditions.

For conditions with 25% of correlations missing, bias magnitude was always below 10% for cells with ks of 50 and 100, except unexpectedly for ρ_{24} with $k = 100$ and $\tilde{n} = 100$. With a k of 25 and $\tilde{n} = 50$, bias was consistently above 10%. With a k of 25 and $\tilde{n} = 100$, the magnitude of the bias decreased below 10% for all ρs except ρ_{14}. In cells with 25% missing data and $k = 25$, bias ranged from 4% to

Table 2: Relative Percentage Bias of Standard Errors of ρ

Study Condition			By ρ						Median
% Missing	k	$\widetilde{n}$	ρ_{12}	ρ_{13}	ρ_{23}	ρ_{14}	ρ_{24}	ρ_{34}	
0	25	50	1.0	1.1	2.2	-3.1	2.0	5.5	1.6
0		100	-1.7	-0.4	4.1	-4.9	0.5	1.1	0.1
0	50	50	-0.1	0.0	0.4	0.7	0.4	0.5	0.4
0		100	-0.6	-2.2	1.4	-2.0	-0.2	-2.0	-1.3
0	100	50	-1.5	-0.4	1.5	-2.3	1.3	-2.0	-1.0
0		100	-1.5	1.4	0.0	-3.2	-1.1	2.5	-0.6
25	25	50	13.7	7.9	11.7	11.1	10.8	11.4	11.3
25		100	7.1	9.3	4.0	10.9	8.2	5.3	7.7
25	50	50	5.2	6.4	4.0	7.1	9.1	8.7	6.8
25		100	6.8	6.1	5.6	3.9	6.6	6.6	6.4
25	100	50	7.3	3.4	7.4	4.6	7.3	4.6	6.0
25		100	5.7	5.7	4.5	5.3	12.2	2.4	5.5
50	25	50	17.3	14.1	12.1	19.9	18.8	17.4	17.4
50		100	16.5	19.0	16.7	12.7	20.0	22.3	17.9
50	50	50	8.7	11.3	5.7	9.0	6.3	10.1	8.9
50		100	11.4	7.3	11.6	10.6	13.6	10.5	11.0
50	100	50	10.6	6.0	9.7	8.3	6.3	10.0	9.0
50		100	9.1	10.6	11.7	7.9	12.4	10.2	10.4

Note: $\widetilde{n}$ represents the average per-study sample size (see equation 19); k = number of studies; % missing = percent of correlations missing in the simulated meta-analysis. Median contains the median relative percentage SE bias by condition.

13.7% with a mean of 9.3% (SD = 2.9). In cells with 50% of correlations missing the bias increased, with larger bias for smaller k. In these cells with a k of 25, bias ranged from 12.1% to 22.3% with a mean of 17.2% (SD = 3.1).

Q-Statistic's Correct Model Rejection Rates

All data were generated with homogeneous correlation matrices. Table 3 shows the proportion of instances that the Q-statistic led to an incorrect inference that there was heterogeneity in the correlation matrices for cells with no missing correlations. Univariate weighting of the Q-statistic led to lower incorrect model rejection rates than did multivariate weighting. However, the average

rejection rate (M = 8.3%, SD = 0.8) for univariate weighting exceeded the nominal level of 5%. The average for multivariate weighting was 17.6% (SD = 5.4). In general, for GLS, the rejection rates increased as k increased. This was not the case for univariate synthesis. For conditions with missing correlations, MI never led to an incorrect rejection of the null hypothesis (i.e., the rejection rate was always 0%), regardless of synthesis method, thus the results are not presented in a table.

Conclusion

Use of MI for meta-analysis resulted in synthesized correlations without substantial parameter estimation bias when data were

Table 3: Homogeneity Assumption Rejection Rates for Omnibus Q-Statistic for Conditions with No Missing Correlations

Study Conditions		Synthesis Method	
k	$\widetilde{n}$	GLS	Univariate
25	50	13.1	8.3
	100	10.9	9.1
50	50	17.2	7.1
	100	16.8	7.6
100	50	25.7	8.8
	100	21.7	9.0

Note: See equation 19 for $\widetilde{n}$. Results for cells with missing correlations are not reported here because MI always resulted in never rejecting the test of homogeneity.

MCAR. Substantial positive standard error bias was found, but typically only for smaller meta-analyses (k of 25) and this bias was higher with larger degrees of missing data. Cheung and Chan (2005) also identified problems with MASEM when values of k were low. Based on the review of applied MASEM studies, however, it appears that most MASEM studies involve larger k values. From these results, use of MI with meta-analysis might be advocated. The results for the test of homogeneity, however, indicate that MI should not be used for testing the homogeneity of correlation matrices across studies. Although MI never resulted in an incorrect inference that there was heterogeneity in the correlation matrices, future research is likely to indicate that when this test should be rejected (i.e., when there is heterogeneity present), MI will have insufficient power.

While parameter and SE estimate results from GLS and univariate weighting utilizing the z-transformation are equivalent when the data being analyzed has no missingness, their results differ for the test of homogeneity. GLS resulted in substantially higher incorrect rejection rates than did the univariate weighting method and the rates for GLS increased as k increased. These findings are consistent with those from Hafdahl (2007) where GLS synthesis resulted in higher

Type I error rates when compared with univariate synthesis when no data were missing. Thus in the context of testing for heterogeneity, the univariate weighting method is recommended.

Limitations and Future Directions

One limitation of this study is the use of only a fixed-effects model both for data generation and model estimation. While this is a limitation, applied MASEM studies most frequently assume a fixed-effects model and therefore the performance of the conditions in this study under this assumption provide an important starting point for this research. A random-effects model, however, might provide a more appropriate fit in many MASEM studies, particularly when important between-study characteristics impact the variability found among studies' correlations. Hafdahl (2008) recently compared GLS with univariate synthesis under a random-effects model with no missing correlations. Hafdahl found that while both methods had high power to reject homogeneity when at least 50 studies were used in the meta-analysis, when fewer than 50 studies were used GLS had far superior power performance over univariate weighting. This difference was particularly noticeable when the average per study sample size was at least 100. Future research should continue to explore the differences between the univariate and multivariate synthesis methods for their power and Type I error control particularly when missing data occur. Given MI's Type I error performance for the test of homogeneity, its performance with between-study heterogeneity should also be evaluated to see if it exhibits the weak power that would be expected.

As noted, MI worked well in most conditions for estimation of correlations and their standard errors (typically substantial standard error bias was only found with 50% of correlations missing). The results for the test of homogeneity with MI, however, seem to indicate a problem with its use. Because Schaefer's (1997) equations for combining the relevant p-values from the χ^2 test of homogeneity seem not to have worked well in the conditions examined here, it would seem that future research should explore whether these

equations can be refined or new equations developed. Additionally, future research should also evaluate the performance of MI with missing data mechanisms other than MCAR.

Use of MASEM techniques will continue to increase as educational researchers use meta-analysis to summarize past research and SEM to investigate relationships between observed and latent variables. It is hoped that the results from this study will help inform the use of, and lead to continued refinement of, MASEM techniques for educational and psychological research.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30-60.

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods & Research*, *28*, 301-309.

Allison, P. D. (2000). COMBCHI [Computer software]. Retrieved September 30, 2005 from University of Pennsylvania website: http://www.ssc.upenn.edu/~allison/#Macros.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*, 545-557.

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*, 243-278. Mahwah, NJ: Lawrence Erlbaum.

Bailey, C. B. (2001). Therapeutic strategies, health status and quality of life among persons with asthma: A meta-analysis. *Dissertation Abstracts International*, *62*, 4B.

Becker, B. J. (1992a). Models of science achievement: Factors affecting male and female performance in school science. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartman, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller. (Eds.), *Meta-analysis for explanation: A casebook.* New York: Russell Sage Foundation.

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

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

Becker, B. J., & Fahrbach, K. (1994, April). *A comparison of approaches to the synthesis of correlation matrices*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

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*, 357-381. New York, NY: Russell Sage Foundation.

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

Brown, S. P., & Stayman, D. M. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research*, *19*, 34-51.

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

Cheung, S. F. (2000). Examining solutions to two practical issues in meta-analysis: Dependent correlations and missing data in correlation matrices. *Dissertation Abstracts International*, *61*, 8B.

Collins, L. M., Schafer, J. L., & Kam, C-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330-351.

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

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

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

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

Fisher, R. A. (1928). *Statistical methods for research workers* (2nd ed.). London: Oliver & Boyd.

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.

Gagné, P., Furlow, C. F., & Beretvas, S. N. (2004). *The equivalence of univariate and Generalized Least Squares methods for meta-analytic synthesis of z-transformed correlations with equal sample size*. Unpublished manuscript.

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples*, 201-218. Hillsdale, NJ: Erlbaum.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. H. Hoyle, (Ed.), *Statistical strategies for small sample research*, 1-29. Thousand Oaks, CA: Sage Publications.

Hafdahl, A. R. (2001). Multivariate meta-analysis for exploratory factor analytic research. *Dissertation Abstracts International*.

Hafdhal, A. R. (2007). Combining correlation matrices: Simulation analysis of improved fixed-effects methods. *Journal of Educational and Behavioral Statistics*, *32*, 180-205.

Hafdahl, A. R. (2008). Combining heterogeneous correlation matrices: Simulation analysis of fixed-effects methods. *Journal of Educational and Behavioral Statistics*, *33*, 507-533.

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

Hershberger, S. L., & Fisher, D. G. (2003). A note on determining the number of imputations for missing data. *Structural Equation Modeling*, *10*, 648-650.

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

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods and Research*, *26*, 329-367.

Insightful Corporation. (2001). S-PLUS 6 for Windows [Computer software]. Seattle, WA.

Klein, H. J., et al. (2001) The assessment of goal commitment: A measurement model meta-analysis. *Organizational Behavior and Human Decision Processes*, *85*, 32-55.

Kubeck, J. E. (2002). Medication adherence: Estimation of a social-cognitive model. *Dissertation Abstracts International*, *62*, 7B.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Manfredo, M. J., Driver, B. L., & Tarrant, M. A. (1996). Measuring leisure motivation: A meta-analysis of the recreation experience preference scales. *Journal of Leisure Research*, *28*, 188-213.

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

Parker, C. P., Baltes, B. B., Young, S. A., Huff, J. W., Altmann, R. A., Lacost, H. A., & Roberts, J. E. (2003). Relationships between psychological climate perceptions and work outcomes: A meta-analytic review. *Journal of Organizational Behavior*, *24*, 389-416.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525-556.

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

Premack, S. L., & Hunter, J. E. (1988). Individual unionization decisions. *Psychological Bulletin*, *103*, 223-234.

Rosenthal, R. (1979). 'The "file drawer problem" and tolerance for null results'. *Psychological Bulletin*, *86*, 638-641.

Rubin, D. B. (1978). Multiple imputations in sample surveys – A phenomenological Bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section*, 20-34. Alexandria, VA: American Statistical Association.

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

SAS Institute. (2005). *SAS/IML software: Changes and enhancements, through release 9.1*. Cary, NC: SAS Institute.

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

Schafer, J. L. (1999). NORM (Version 2.03 for Windows) [Computer software]. Retrieved April 1, 2005, from http://www.stat.psu.edu/~jls/misoftwa.html.

Schafer, J. L., & Graham, J. W. (1999). On the performance of multiple imputation for multivariate data with small sample sizes. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research*, 1-29. Mahwah, NJ: Lawrence Erlbaum.

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 (4), 545-571.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*, 317-329.

SPSS. (2006). *Statistical Package for the Social Sciences*. Chicago, IL: SPSS.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251.

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology*, *46*, 259-293.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*, 231-249.