INVESTIGATING RELIABILITIES FOR MULTI-LEVEL DATA WITH MISSING VALUES

by

MINJU HONG

(Under the Direction of Zhenqiu Laura Lu)

ABSTRACT

Reliability indicates the internal consistency of a test. In educational studies, reliability is a key feature for a test. Researchers have proposed many traditional reliability estimates, such as coefficient alpha and coefficient omega. However, traditional reliability indices do not deal with the data hierarchy, even though the multilevel structure is prevalent in most of the educational data sets. And then level-specific reliabilities (i.e., the within-group and between-group level) for multilevel data structure have been recently proposed. But the new approach has not considered the influence of missing data, which is very common in educational studies, social and behavioral areas.

The aim of this study is to investigate the reliability estimation for multi-level data with missing values. We first reviewed the traditional single level reliabilities and the level-specific reliabilities. And then we proposed a new model, the multilevel confirmatory factor with missing values. This model was illustrated by analyzing a real data set, the 2018 PISA data. We found some differences between the single-level and the level-specific reliability estimates. And then we further investigated the performance of the model by conducting simulation studies under various conditions. We focused on a two-levels single-factor model with six indicators. Six reliability estimates (single-level coefficient alpha α, within-group coefficient alpha α_W, between-group coefficient alpha α_B, single-level coefficient omega ω, within-group coefficient omega ω_W, and between-group coefficient omega ω_B) were investigated and compared. Various simulations conditions were considered, number of clusters, cluster sizes, intraclass correlation (ICCs), missing data mechanisms, missing data proportions, and missing data techniques. Parameter biases and convergence rates were compared.

Results showed that (1) missing values have negative impacts on the reliability estimation and with the proportion of missing data increased, the percentage bias increased while the convergence rates decreased, (2) listwise deletion method performed worse than the full information maximum likelihood (FIML) method, (3) ICC is importance in the reliability estimation and when ICC was large (say, .30 in our simulation), the between-group level reliability estimates outperformed the within-group reliabilities, (4), the coefficient omega performed better than the coefficient alpha if the tau-equivalent assumption is violated. Limitations of the current model and future research were also discussed in the last part.

INDEX WORDS: Reliability, Multilevel Confirmatory Factor Analysis, Missing Data

INVESTIGATING RELIABILITIES FOR MULTI-LEVEL DATA WITH MISSING VALUES

by

MINJU HONG

B.A., Korea University, Republic of Korea, 2012

M.A., Korea University, Republic of Korea, 2015

M.S., University of Georgia, 2020

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

INVESTIGATING RELIABILITIES FOR MULTI-LEVEL DATA WITH MISSING VALUES

by

MINJU HONG

Major Professor: Zhenqiu Laura Lu
Committee: Allan S. Cohen
Seock-Ho Kim
Justin Strait

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2022

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Reliability is one of the most important features of a measurement tool such as a test or a questionnaire. A high reliability ensures the consistency of the test results (Feldt & Brennan, 1989) and the necessary conditions for the test's validity (Messick, 1989). A test in social and behavioral sciences usually requires reporting the reliability (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014; Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999). In the early of 1940s, reliability was defined as a correlation coefficient of the observed scores (Kelly, 1942) or the ratio of the true score variance to the total score variance (Kaitz, 1945). According to Crocker and Algina (1986), if we administer the same test twice or administer two similar tests, we could then use the coefficient of equivalence (derived from the alternative form method) or the coefficient of stability (derived from the test-retest method) as reliability measures. However, as pointed out by Cronbach (1943), two or more test administrations cause bias in the reliability estimates, so we use the observed scores obtained from a single test administration. In this case, the reliability coefficient is calculated by the Spearman-Brown prophecy formula based on the split-half method or the internal consistency reliability measures such as KR 20 or KR 21 calculated by the Kuder Richardson formula (Kuder & Richardson, 1937), the coefficients calculated by the Hoyt formula (Hoyt, 1941), or the coefficient alpha (Cronbach, 1951).

However, reliability measures based on the classical test theory had a limitation because the true scores and the errors could not be observed directly (Miller, 1995). By using factor models (Jöreskog, 1971), we can estimate reliabilities by treating the true scores and the error as the latent variables. In structural equation modeling (SEM) or factor analysis (FA), factors are latent or unobserved. For example, in educational area, the true ability of the examinees or the underlying construct that the test targets to measure are factors and their indicators are the test items.

Under the factor analysis framework, there are three types of measurement models to explain the relationship between factors and indicators, parallel, tau-equivalent and congeneric. the If the test is parallel, it assumes the following: (a) the same unit of measurement, and (b) the same precision of measurement. That is, assumption (a) means that the change of the factor score causes the same magnitude of change in the scores of all indicators (i.e., the same factor loadings of all indicators in the model), and assumption (b) means that the variability of the errors of all indicators' scores are the same (i.e., the same error variances). When we hold assumption (a) but relax assumption (b), the test is tau-equivalent. If we relax both assumptions, the test is congeneric.

Because two assumptions for the parallel model are too strict in practical testing situations (Crocker & Algina, 1986), reliability measures have been proposed based on either the tau-equivalent or the congeneric model. The most common reliability estimate is the coefficient alpha (Cronbach, 1951, 2004). As reviewed in the literature (Deng & Chan, 2017; Dunn et al., 2014; Hogan et al., 2000; McNeish, 2018; Yang, & Green, 2011), the reasons the coefficient alpha is widely reported in most social science studies are as follows: (a) the coefficient alpha has long been accepted as the best-known reliability index (i.e., the coefficient alpha is familiar to most of the researchers of the social science studies); (b) The coefficient alpha has a much simpler calculation process than the alternative reliability estimates such as the coefficient omega (McDonald, 1978), the coefficient beta (Revelle, 1979), or greatest lower bound reliability (Bentler, 1972; Jackson & Agunwamba, 1977; Woodhouse & Jackson, 1977); (c) the coefficient alpha is based on the fully saturated model so that the model-fit is not a concern; and (d) most computer package programs (e.g., SPSS, Mplus, SAS, etc.) set the coefficient alpha as a default.

In spite of its popularity, it is controversial to use the coefficient alpha because the tau-equivalent assumption could be violated in actual test administration situations such as the different factor loadings estimated from the factor model (Bentler, 2009; Cortina, 1993; Feldt & Qualls, 1996; Green & Yang, 2009a, 2009b; Raykov, 1997a; Revelle & Zinbarg, 2009; Sijtsma, 2009; Yang, & Green, 2011). In particular, for a one-factor model, if the differences among the factor loadings are large (e.g., above .6), the coefficient omega should be used as an alternative to the coefficient alpha (Dunn et al., 2014; Edwards et al., 2021; Raykov & Marcoulides, 2015; Zhang & Yuan, 2016; Zinbarg et al., 2005).

However, as test administration has become more complicated, studies reported that the traditional reliability estimates (e.g., the coefficient alpha or coefficient omega) may be biased (Geldhof et al., 2014). One problem comes from the multilevel structure of the collected data (Bonito et al., 2012; Geldhof et al., 2014; Jak & Jorgensen, 2017; Jeon et al., 2009; Sideridis et al., 2018). For example, researchers collect data using multistage sampling. If they sample schools in the first stage and in the second stage they collect data from the students of the selected schools, then the data have two levels with the student level nested within the school level. In this case, the students from the same school share the common school features, so there is a dependency among students' data, which violates the assumption of independent residuals.

As such, multilevel data may lead to biased reliability estimates (e.g., Snijders & Bosker, 1999). Geldhof et al. (2014) found that if the true reliability values at both levels are different, then the traditional single-level reliability estimates are inaccurate. Also, Bonito et al. (2012) demonstrated that the single-level reliability measures could be the average of the reliability estimates of both levels.

Another problem comes from the missing values in a data set. For example, examinees may skip answering some test items because of a lack of knowledge or a time limit (Çokluk and Kayri, 2011). If we do not consider these missing values, the reliability estimates would be biased because the data with missing values would not represent the features of the population. Some factor of missing values—for example, the missing data proportions, the missing data mechanisms and the missing data techniques—play significant roles in estimating reliability measures (Cuesta Izquierdo & Fonseca Pedrero, 2014; Downey & King, 1998; Enders, 2003, 2004; McDonald et al., 2000; Raykov, 2009b).

The limitations of the studies above are as follows: (a) The studies using the multilevel reliability indices (Bonito et al., 2012; Geldhof et al., 2014; Jak & Jorgensen, 2017; Jeon et al., 2009; Sideridis et al., 2018) did not consider the effect of the missing values in the reliability estimation, and (b) the studies about the effects of the missing values on the estimation of the traditional reliability indices (Cuesta Izquierdo & Fonseca Pedrero, 2014; Downey & King, 1998; Enders, 2003, 2004; McDonald et al., 2000; Raykov, 2009b) did not consider the multilevel data structure.

Therefore, this study aims to propose a new model to address these problems in estimating reliabilities and to investigate the performance of both traditional single-level and multilevel (i.e., within-group level and between-group level) reliability estimates. The simulation study considers the conditions of missing values (i.e., missing data mechanisms, missing data proportions, and missing data techniques) and the conditions of multilevel data structure (i.e., number of clusters, cluster sizes, and intra-class correlations).

The following sections contain a literature review on single-level and multilevel reliability estimates and the impact of missing values on reliability estimates. Next, the empirical data analysis of this study shows the performance of single-level and level-specific reliability estimates for actual educational data. By conducting a simulation study, the performance of the reliability estimates is compared regarding the conditions of missing values and multilevel data structure. Implications and further directions of this study are mentioned in the last section.

CHAPTER 2

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

Reliability for Single-Level Data

Based on the classical test theory (CTT), Spearman (1904) proposed a classical true score model to explain the components of the observed score related to the true score and the error. Suppose there is a test containing I items. the reliability is defined as the coefficient of the internal consistency of the items in the test (Allen & Yen, 1979). Because the test includes the multiple items from X_1 to X_I, the composite of the observed scores of all items in the test is $X = \sum_{i=1}^{I} X_i$ and it is decomposed as

$$X = T + E \tag{1}$$

where $T = \sum_{i=1}^{I} t_i$ is the true score and $E = \sum_{i=1}^{I} e_i$ is the error, assuming that (a) the covariance between the true score and error is zero ($\sigma_{TE} = 0$), (b) the mean of the errors is equal to zero ($\mu_E = 0$), and (c) the covariances of the pairs of errors are zero ($\sigma_{E_i E_j} = 0$ for $i \neq j \in 1, \ldots, I$). Then, the observed score variance σ_X^2 is decomposed by

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \tag{2}$$

where σ_T^2 is the true score variance, and σ_E^2 is the error variance. Then, reliability is defined as the ratio of the true score variances over the observed score variances (Crocker & Algina, 1986)

$$\rho_{XT} = \frac{\sigma_T^2}{\sigma_X^2} \tag{3}$$

The denominator of Equ. (3), the observed score variance σ_X^2, can be directly calculated because of $X = \sum_{i=1}^{I} X_i$.

$$\sigma_X^2 = Var(X) = Var\left(\sum_{i=1}^{I} X_i\right) = \sum_{i=1}^{I} Var(X_i) + 2\sum_{i\ddagger j} Cov(X_i X_j)$$
$$= \sum_{i=1}^{I} \sigma_{X_1}^2 + 2\sum_{i\ddagger j} \sigma_{X_i X_j} \quad (4)$$

It is the sum of all elements in the variance–covariance matrix of the observed scores (i.e., Matrix 1) as follows,

$$\begin{array}{c|ccccc|}
 & X_1 & X_2 & \cdots & X_{I-1} & X_I \\ \hline
X_1 & \sigma_{X_1}^2 & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_{I-1}} & \sigma_{X_1X_I} \\
X_2 & \sigma_{X_2X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2X_{I-1}} & \sigma_{X_2X_I} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
X_{I-1} & \sigma_{X_{I-1}X_1} & \sigma_{X_{I-1}X_2} & \cdots & \sigma_{X_{I-1}}^2 & \sigma_{X_{I-1}X_I} \\
X_I & \sigma_{X_IX_1} & \sigma_{X_IX_2} & \cdots & \sigma_{X_IX_{I-1}} & \sigma_{X_I}^2 \\ \hline
\end{array}$$

where $i = 1, 2, \ldots, I$ is the number of items, $\sigma_{X_i}^2$ is the observed score variance of the item i, and $\sigma_{X_iX_j}$ is the covariance of the observed scores of the items i and j for $i \neq j$.

However, the numerator of the reliability of Equation (3), the true score variance σ_T^2, cannot be simply obtained because the true score cannot be directly measured. We use factor models to estimate the true score (Bollen, 1989; Raykov, 1997b). Suppose the test is designed to measure M factors (i.e., constructs underlying the test) by I indicators (i.e., items). M factors are the latent constructs targeted to be measured by the test. For example, if the test is designed to examine students' understandings of multiple mathematical operations—such as addition, subtraction, multiplication, and division—each of the factors indicates the knowledge of each operation. With factor models, the observed score of the test X is decomposed by

$$X = \sum_{m=1}^{M}\sum_{i=1}^{I} \lambda_{im}\eta_m + \sum_{i=1}^{I} \varepsilon_i \quad (5)$$

where X is the observed score, λ_{im} is the factor loading of the mth factor on the ith item, η_m is the factor score of the mth factor, and ε_i is the error of the ith test item (Brown, 2015; Mulaik, 2009). Here, we assume the factor scores are normally distributed satisfying the mean is equal to zero and the variance is equal to 1 ($\eta_m \sim N(0,1)$) and the errors are normally distributed satisfying the mean is equal to zero and the variance θ_{ii} for the ith test item ($\varepsilon_i \sim N(0, \theta_{ii})$). In addition, the covariance between the factor and errors are zero ($\sigma_{\eta_m \varepsilon_i} = 0$), and the covariance among the errors are zero ($\sigma_{\varepsilon_i \varepsilon_j} = 0$ for $i \neq j$).

For simplicity, suppose the items in the test aim to measure one common factor; thus, the model is a single factor model. In this case, Equation (5) is simplified to

$$X = \sum_{i=1}^{I} \lambda_i \eta + \sum_{i=1}^{I} \varepsilon_i \tag{6}$$

where the factor score is $\eta_m \sim N(0,1)$. With this model, the observed score variance is $\sigma_{X_i}^2 = Var(X_i) = Var(\lambda_i \eta + \varepsilon_i) = Var(\lambda_i \eta) + Var(\varepsilon_i) + 2Cov(\lambda_i \eta, \varepsilon_i)$. Additionally, we assume that $Cov(\lambda_i \eta, \varepsilon_i) = 0$, then $\sigma_{X_i}^2 = Var(\lambda_i \eta) + Var(\varepsilon_i) = \lambda_i^2 Var(\eta) + Var(\varepsilon_i)$. Also, we assume that $Var(\eta) = 1$ and $Var(\varepsilon_i) = \theta_{ii}$, thus, $\sigma_{X_i}^2 = \lambda_i^2 + \theta_{ii}$.

In addition to the observed score variance, the covariance of the observed scores is $\sigma_{X_i X_j} = Cov(X_i, X_j) = Cov(\lambda_i \eta + \varepsilon_i, \lambda_j \eta + \varepsilon_j) = \lambda_i \lambda_j Cov(\eta + \varepsilon_i, \eta + \varepsilon_j)$. Because $Cov(\eta + \varepsilon_i, \eta + \varepsilon_j) = Var(\eta) + Cov(\eta, \varepsilon_i) + Cov(\eta, \varepsilon_j) + Cov(\varepsilon_i, \varepsilon_j)$ and $Cov(\eta, \varepsilon_i) = Cov(\eta, \varepsilon_j) = Cov(\varepsilon_i, \varepsilon_j) = 0$, then, $\sigma_{X_i X_j} = \lambda_i \lambda_j Var(\eta)$. And $Var(\eta) = 1$, thus, $\sigma_{X_i X_j} = \lambda_i \lambda_j$. From these, we can rewrite Matrix 1 to Matrix 2 as the following:

$$\begin{array}{c|ccccc|}
 & X_1 & X_2 & \cdots & X_{I-1} & X_I \\
\hline
X_1 & \lambda_1^2 + \theta_{11} & \lambda_1 \lambda_2 & \cdots & \lambda_1 \lambda_{I-1} & \lambda_1 \lambda_I \\
X_2 & \lambda_2 \lambda_1 & \lambda_2^2 + \theta_{22} & \cdots & \lambda_2 \lambda_{I-1} & \lambda_2 \lambda_I \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
X_{I-1} & \lambda_{I-1} \lambda_1 & \lambda_{I-1} \lambda_2 & \cdots & \lambda_{I-1}^2 + \theta_{I-1I-1} & \lambda_{I-1} \lambda_I \\
X_I & \lambda_I \lambda_1 & \lambda_I \lambda_2 & \cdots & \lambda_I \lambda_{I-1} & \lambda_I^2 + \theta_{II} \\
\hline
\end{array}$$

From this variance–covariance matrix, the elements in the diagonal are the observed score variances, and the elements in the off-diagonal are the covariances of the pair of the two observed scores. Then, Matrix 2 is divided into two matrices including a matrix of the true score variances (i.e., Matrix 3) and a matrix of the error variances (i.e., Matrix 4) as follows:

$$\begin{array}{c|ccccc|} & X_1 & X_2 & \dots & X_{I-1} & X_I \\ \hline X_1 & \lambda_1^2 & \lambda_1\lambda_2 & \dots & \lambda_1\lambda_{I-1} & \lambda_1\lambda_I \\ X_2 & \lambda_2\lambda_1 & \lambda_2^2 & \dots & \lambda_2\lambda_{I-1} & \lambda_2\lambda_I \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{I-1} & \lambda_{I-1}\lambda_1 & \lambda_{I-1}\lambda_2 & \dots & \lambda_{I-1}^2 & \lambda_{I-1}\lambda_I \\ X_I & \lambda_I\lambda_1 & \lambda_I\lambda_2 & \dots & \lambda_I\lambda_{I-1} & \lambda_I^2 \\ \hline \end{array} + \begin{array}{c|ccccc|} & X_1 & X_2 & \dots & X_{I-1} & X_I \\ \hline X_1 & \theta_{11} & 0 & \dots & 0 & 0 \\ X_2 & 0 & \theta_{11} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{I-1} & 0 & 0 & \dots & \theta_{I-1I-1} & 0 \\ X_I & 0 & 0 & \dots & 0 & \theta_{II} \\ \hline \end{array}$$

Because the composite of the observed scores is $X = \sum_{i=1}^{I} X_i$, the composite of the true scores is $T = \sum_{i=1}^{I} T_i$, and the composite of the errors is $E = \sum_{i=1}^{I} E_i$, we rewrite Equation (2) to Equation (7) as the following:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 = \sum_{i=1}^{I} \lambda_i^2 + 2\sum_{i \neq j} \lambda_i\lambda_j + \sum_{i=1}^{I} \theta_{ii} = \left(\sum_{i=1}^{I} \lambda_i\right)^2 + \sum_{i=1}^{I} \theta_{ii} \tag{7}$$

There are three types of measurement models explain the relationship between the factor and the indicators (Jöreskog, 1971), parallel measurement, tau equivalent, and congeneric. Parallel measurements assume (a) the same factor loadings, $\lambda_1 = \lambda_2 = \cdots = \lambda_I = \lambda$, and (b) the same error variances of the indicators, $\theta_{11} = \theta_{22} = \cdots = \theta_{II} = \theta$. However, in the actual testing situation, it is difficult to develop a test that meets two assumptions (Crocker & Algina, 1986; Novick & Lewis,1966). If we assume (a) without (b), the measurement is tau equivalent. If we relax both assumptions, the measurement is congeneric. If the measurement is parallel, then the variance–covariance matrix of the parallel model (i.e., Matrix 5) is as follows:

$$\begin{array}{c|ccccc|} & X_1 & X_2 & \dots & X_{I-1} & X_I \\ \hline X_1 & \lambda^2+\theta & \lambda^2 & \dots & \lambda^2 & \lambda^2 \\ X_2 & \lambda^2 & \lambda^2+\theta & \dots & \lambda^2 & \lambda^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{I-1} & \lambda^2 & \lambda^2 & \dots & \lambda^2+\theta & \lambda^2 \end{array}$$

$$\begin{array}{c|ccccc|}
X_I & \lambda^2 & \lambda^2 & \dots & \lambda^2 & \lambda^2+\theta \\ \hline
\end{array}$$

If the measurement is tau equivalent, then the variance–covariance matrix of the tau equivalent model (i.e., Matrix 6) is as follows:

$$\begin{array}{c|ccccc|}
 & X_1 & X_2 & \dots & X_{I-1} & X_I \\ \hline
X_1 & \lambda^2+\theta_{11} & \lambda^2 & \dots & \lambda^2 & \lambda^2 \\
X_2 & \lambda^2 & \lambda^2+\theta_{22} & \dots & \lambda^2 & \lambda^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_{I-1} & \lambda^2 & \lambda^2 & \dots & \lambda^2+\theta_{I-1I-1} & \lambda^2 \\
X_I & \lambda^2 & \lambda^2 & \dots & \lambda^2 & \lambda^2+\theta_{II} \\ \hline
\end{array}$$

If the measurement is congeneric, the variance–covariance matrix of the congeneric model is Matrix 2 shown as follows:

$$\begin{array}{c|ccccc|}
 & X_1 & X_2 & \dots & X_{I-1} & X_I \\ \hline
X_1 & \lambda_1^2+\theta_{11} & \lambda_1\lambda_2 & \dots & \lambda_1\lambda_{I-1} & \lambda_1\lambda_I \\
X_2 & \lambda_2\lambda_1 & \lambda_2^2+\theta_{22} & \dots & \lambda_2\lambda_{I-1} & \lambda_2\lambda_I \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
X_{I-1} & \lambda_{I-1}\lambda_1 & \lambda_{I-1}\lambda_2 & \dots & \lambda_{I-1}^2+\theta_{I-1I-1} & \lambda_{I-1}\lambda_I \\
X_I & \lambda_I\lambda_1 & \lambda_I\lambda_2 & \dots & \lambda_I\lambda_{I-1} & \lambda_I^2+\theta_{II} \\ \hline
\end{array}$$

The well-known reliability Cronbach's alpha (coefficient alpha α; proposed by Cronbach, 1951) is based on parallel or tau equivalent measurement. It is calculated by

$$\alpha = \frac{I^2\bar{\sigma}_{ij}}{\sum \sigma_X^2} \tag{8}$$

where I is the total number of items, $\bar{\sigma}_{ij}$ ($i \neq j$ with $i, j = 1, \dots, I$) is the average of off-diagonal elements of the observed scores' variance–covariance matrix (i.e., Matrix 6), and $\sum \sigma_X^2$ is a sum of all elements of the observed scores' variance–covariance matrix (i.e., Matrix 6). The reliability coefficient omega (McDonald, 1978) is for the congeneric measurement and is calculated by

$$\omega = \frac{Var(\sum_{i=1}^{I}\lambda_i\eta)}{Var(\sum_{i=1}^{I}\lambda_i\eta + \sum_{i=1}^{I}\varepsilon_i)} = \frac{(\sum_{i=1}^{I}\lambda_i)^2}{(\sum_{i=1}^{I}\lambda_i)^2 + \sum_{i=1}^{I}\theta_{ii}} \tag{9}$$

where $i = 1, \ldots, I$ indicates each of items in the test, $Var(\sum_{i=1}^{I} \lambda_i \eta)$ is the true score variance, $Var(\sum_{i=1}^{I} \lambda_i \eta + \sum_{i=1}^{I} \varepsilon_i)$ is the observed score variance, λ_i is the factor loading, and θ_{ii} is the error variance of the ith test item.

The coefficient alpha is popular and convention (Hogan et al., 2000) because it is easily calculated from the observed scores directly; in other words, we obtain the coefficient alpha from the fully saturated model. Thus, we do not need to consider the model fit as the other reliability measures should be considered under the factor analysis framework and most computer programs provide the coefficient alpha as the default (McNeish, 2018).

However, much of the literature has pointed out the limits of the coefficient alpha (Cortina, 1993; Green et al., 1977; Green & Yang, 2009a; Raykov, 1997a, 1997b; Sijtsma, 2009; Yang & Green, 2011; Zimmermanet al., 1993) because it is unbiased only when the data meet the parallel or tau-equivalent assumption, in which all factor loadings are the same. However, it is not a practical assumption in many social sciences when developing a measurement tool (e.g., test items in an exam or questionnaire). Hence, other reliability measures have been proposed, such as the coefficient omega (McDonald, 1978), the coefficient beta (Revelle, 1979), and greatest lower bound reliability (Bentler, 1972; Jackson & Agunwamba, 1977; Woodhouse & Jackson, 1977).

Among them, the coefficient omega has been often investigated not only in empirical data analyses, but also in simulation studies (Deng & Chan, 2017, Duun et al., 2014; Edward et al., 2021; Zinbarg, et al., 2005). These investigations have shown that the coefficient omega is more robust than the coefficient alpha for estimating the reliability of data sets which are not satisfied by the tau-equivalent assumption with a large difference in the factor loadings; an example of this is a case in which the difference between the average factor loadings and the individual factor loading are above .20, as examined in Raykov and Marcoulides (2015) or when the difference between the highest (.80) and the lowest (.20) factor loadings is .60, as investigated in Green and Yang (2009a), even though Edward et al. (2021) showed that the coefficient alpha and the coefficient omega performed well with small sample sizes (e.g., 30 samples) and low population reliability values (e.g., .20 for the data simulation). However, the model estimating the coefficient omega showed more failures in estimation (i.e., the non-convergence issues) when the sample sizes are small, 250 samples, and the population reliability values are small, around .30 (Hancock & An, 2020). In addition, if there is a model misspecification (i.e., the model has a poor fit), then the coefficient omega is biased (Zinbarg et al., 2005).

Reliability for Multi-Level Data

The data in the field of education often have multilevel (Bryk and Raudenbush, 1987). For example, when a researcher collects data from students and their teachers simultaneously, the data have a two-level hierarchy, including the student level and the teacher level. Also, suppose the student is administered the achievement test once a year; then, the data have two levels (the test score level and the student level) because the annually measured scores belong to each student.

The lower level of the data (e.g., the student level for the first example and the test score level for the second example) is referred to as the within-group level (or level-1), and the upper level of the data (e.g., the teacher level for the first example, and the student level for the second example) is referred to as the between-group level (or level-2). However, these two-level data cannot satisfy the assumption of traditional data analysis methods because the observed values at the within-group level share the common feature of a between-group level unit (Raudenbush &

Bryk, 2002). For the first example, the students who belong to the same class share the features of the teacher. For the second example, the characteristics of the student affect the annually measured test scores. In these cases, the assumption is that the covariances among the errors (under the classical true score model, $\sigma_{E_i E_j} = 0$ for $i \neq j$; under the factor analysis model, $\sigma_{\varepsilon_i \varepsilon_j} = 0$ for $i \neq j$) is violated for analyzing a two-level data.

Under the factor analysis model framework, Muthén (1994, 1997) applied the multilevel confirmatory factor analysis (MCFA) model to analyze the educational data set.

From Figure 1, suppose a test has I test items, and there are K examinees belonging to G schools (i.e., groups). For simplicity, we assume it is a single factor model. Then, we can specify the two-level MCFA model to decompose the observed score $X_{kg} = \sum_{i=1}^{I} X_{ikg}$ and

$$X_{ikg} = X_{Big} + X_{wikg} = \left(\lambda_{Bi}\eta_{Bg} + \varepsilon_{Big}\right) + \left(\lambda_{wi}\eta_{wkg} + \varepsilon_{wikg}\right) \tag{10}$$

where X_{Big} is the component of X_{ikg} at the between-group level, and X_{wikg} is the component of X_{ikg} at the within-group level. At the between-group level, X_{Big} is the component of the observed score X at the between-group level decomposed by the factor loading of the ith test item (λ_{Bi}), the factor score of the kth examinee in the gth group (η_{Bg}), and the error of the ith test item for the kth examinee in the gth group (ε_{Big}). At the within-group level, X_{wikg} is the component of the observed score at the within-group level decomposed by the factor loading of the ith test item (λ_{wi}), the factor score of the kth examinee in the gth group (η_{wkg}), and the error of the ith test item for the kth examinee in the gth group (ε_{wikg}).

For the factor scores and errors at both levels, we assume that (a) the factor scores are normally distributed with the mean as zero and the variance as 1 (i.e., $\eta_{Bg} \sim MVN(0, 1)$ and $\eta_{wkg} \sim MVN(0,1)$), (b) the errors are normally distributed with the mean as zero and the variance as θ_{wii} for the within-group level and θ_{Bii} for the between-group level (i.e., $\varepsilon_{Big} \sim MVN(0, \theta_{Bii})$ and $\varepsilon_{wikg} \sim MVN(0, \theta_{wii})$), (c) the covariances between the factor score and the errors are equal to zero (i.e., $\sigma_{\eta_{Bg}\varepsilon_{Big}} = 0$ and $\sigma_{\eta_{wkg}\varepsilon_{wikg}} = 0$), and (d) the covariances among the errors are equal to zero (i.e., $\sigma_{\varepsilon_{Big}\varepsilon_{Bjg}} = 0$ and $\sigma_{\varepsilon_{wikg}\varepsilon_{wjkg}} = 0$ for $i \neq j$).

Under the MCFA model, Geldhof et al. (2014) investigated the level-specific reliability estimates; their idea was to apply coefficients alpha and omega in Equations (8) and (9) to estimate the reliability measures within each level of the data separately. Suppose a test aims to measure a single factor. If the test is tau equivalent, at the within-group level, by assuming that factor loadings are $\lambda_{w1} = \lambda_{w2} = \cdots \lambda_{wi} = \cdots = \lambda_{wI} = \lambda_w$, error variances θ_{wii} are varied among the ith test item, and at the between-group level, factor loadings are $\lambda_{B1} = \lambda_{B2} = \cdots \lambda_{Bi} = \cdots = \lambda_{BI} = \lambda_B$, and error variances θ_{Bii} are varied for the ith test item, the test is tau equivalent. Then, level-specific alpha, the within-group level coefficient alpha α_w, and the between-group level coefficient alpha α_B are calculated by

$$\alpha_w = \frac{I^2 \bar{\sigma}_{wij}}{\sum \sigma^2_{X_w}} \tag{11}$$

$$\alpha_B = \frac{I^2 \bar{\sigma}_{Bij}}{\sum \sigma^2_{X_B}} \tag{12}$$

where I is the total number of items, $\bar{\sigma}_{wij}$ ($i \neq j$ with $i, j = 1, \dots, I$) is the average of off-diagonal elements of the observed scores' variance–covariance matrix at the within-group level, $\sum \sigma^2_{X_w}$ is a sum of all elements of the observed scores' variance–covariance matrix at the within-group level, $\bar{\sigma}_{Bij}$ ($i \neq j$ with $i, j = 1, \dots, I$) is the average of off-diagonal elements of the observed scores' variance–covariance matrix at the between-group level, and $\sum \sigma^2_{X_B}$ is a sum of all elements of the observed scores' variance–covariance matrix at the between-group level.

If the test is congeneric, that is, when we relax the assumptions of the same factor loadings at the within-group and between-group levels, then level-specific omega, that is the within-group level coefficient omega ω_w and the between-group level coefficient omega ω_B, are calculated by

$$\omega_w = \frac{(\sum_{i=1}^{I} \lambda_{wi})^2}{(\sum_{i=1}^{I} \lambda_{wi})^2 + \sum_{i=1}^{I} \theta_{wii}} \tag{13}$$

$$\omega_B = \frac{(\sum_{i=1}^{I} \lambda_{Bi})^2}{(\sum_{i=1}^{I} \lambda_{Bi})^2 + \sum_{i=1}^{I} \theta_{Bii}} \tag{14}$$

where $i = 1, \dots, I$ indicates each of the items in the test, λ_{wi} is the factor loading at the within-group level, θ_{wii} is the error variance of the ith test item at the within-group level, λ_{Bi} is the factor loading at the between-group level, and θ_{ii} is the error variance of the ith test item at the between-group level.

Instead of reporting level-specific reliability, which is important, the studies of multilevel modeling have rarely reported the reliability information (Kim et al., 2016). In reviewing the literature about application of level-specific reliability estimates, the results are as follows. Based on the empirical data analysis (Jak & Jorgensen, 2017; Jeon et al., 2009; Sideridis et al., 2018), when we ignore the data hierarchy, we obtain misleading reliability estimates because the results are the aggregated value (e.g., the average or weighted sum) of the reliabilities at the within- and between-group level. Ignoring the multilevel data structure distorted not only the results of the reliability estimation but also the other statistical analyses, such as the factor analysis (Kim et al., 2016; Padgett & Morgan, 2021; Pornprasertmanit at el., 2014), the structural equation modeling (Hsu et al., 2007; Ryu & West, 2009), and the cluster randomized study (Cho & Preacher, 2016).

By conducting the simulation study, Bonito et al. (2012) investigated the performance of the individual- and group-level reliability estimates based on the two-level multilevel model

(MLM) (Raudenbush & Bryk, 2002). Bonito et al. (2012) found that under the condition of five items in the test, the reliability ignoring the data hierarchy showed more inflated values than the individual- and group-level reliability estimates did. However, the difference between the reliability estimates with or without considering the two-level data structure was small under the condition of ten items in the test and all group sizes (e.g., two, four, or six individuals in each group) in this study.

To extend the results of Bonito et al. (2012) and Geldhof et al. (2014) showed the necessity of using the level-specific reliability based on a multilevel confirmatory factor analysis (MCFA) model with two factors and six indicators under a two-level data structure. Specifically, they manipulated the sample sizes regarding two-level data structure (e.g., the number of clusters as 50, 100, 200, and the cluster sizes as 2, 15, 30), the intra-class correlations (ICCs .05, .25, .50, .75), and the population values of the reliabilities (.85 as high, .37 as low for α, ω, and H). For the population values of the reliabilities, Geldhof et al. (2014) manipulated the factor loading differences; for the high population values of the reliabilities, they varied the factor loadings from .60 to .80 (i.e., congeneric condition), while for the low population values, they fixed the factor loadings at .30 (i.e., tau-equivalent condition).

Geldhof et al. (2014) found that except H (because H showed the worst results under all simulated conditions), under the conditions of the high ICCs (over .05), the level-specific and especially the between-group level reliability indices showed less bias than the single-level or the within-group level reliability measures. In addition, under the conditions of the high population values of the reliability (i.e., the congeneric; the violation of the tau-equivalent condition), the between-group level ω outperformed the between-group level α. However, for the larger sample sizes (i.e., 6,000 samples with 30 individuals per cluster and 200 clusters), the estimates of the coefficient ω reported more biased results than the coefficient α. These results were the extended findings of the studies comparing the estimates of the coefficient α and the coefficient ω at the single level (Edward et al., 2021; Green & Yang, 2009a; Hancock & An, 2020; Raykov & Marcoulides, 2015; Zinbarg et al., 2005).

However, instead of the advantages of the level-specific reliability estimates, Lai (2021) pointed out the possibility of inflated values of these estimates under certain data conditions. For example, if the test is not unidimensional but multidimensional, and the underlying constructs are overlapped (e.g., the test is designed to measure the interdisciplinary understandings of the science subjects, such as biology, physics, chemistry, and earth science), then the level-specific coefficient ω would be overestimated.

Reliability with Missing Data

Missing Data and Missing Data Mechanisms

Missing data is the not observed values in the data set because of many reasons (Little & Rubin, 2002; McKnight et al., 2007; Rubin, 1976). When the examinee answers the questionnaire and these answers are recorded into the data set, the information provided by the examinee becomes the observed data. However, when the examinee's answers are not recorded for some reason (e.g., the examinee refused to answer, skipped some questions, or failed to answer all questions within the permitted time, or the answers were mistakenly omitted from the data entry), then the data are missing.

The literature has defined the missing data mechanism as the process leading to the missingness (Rubin, 1976), the reasons for the missing data (Little & Rubin, 2002), the assumptions of probabilistic explanations about the missing data (Peugh & Enders, 2004), or the probability of the missing data given the information about the entire data set (McKnight et al., 2007). In sum, the missing data mechanism is the statistical relationship among the observed and the missing data in the entire data set.

Suppose there is a data set of X_{ikg}, consisting of the scores of the kth ($k = 1, 2, \dots, K$) examinee within the gth ($g = 1, 2, \dots, G$) group on the ith ($i = 1, 2, \dots, I$) test item. Because the data set has the missing values, the missing data indicator of the corresponding score X_{ikg} is specified as M_{ikg}: if X_{ikg} is missing, M_{ikg} is one, or if X_{ikg} is observed, then, M_{ikg} is zero.

Little and Rubin (2002) explained three missing data mechanisms. First, if the probability of missing data is not dependent on any factors of the entire data set, it is missing completely at random (MCAR). Under MCAR, ϕ_{ikg}, the probability of missingness on X_{ikg}, is specified as

$$\phi_{ikg} = p\left(M_{ikg} = 1 \middle| X_{ikg}, \phi\right) = p\left(M_{ikg} = 1 \middle| X_{obs}, X_{miss}, \phi\right) = p\left(M_{ikg} = 1 \middle| \phi\right) \qquad (15)$$

where X_{obs} is a matrix of the observed values in the data set of the score X_{ikg}, X_{miss} is a matrix of the missing values in the data set of the score X_{ikg}, and ϕ is a set of parameters for the missing data distribution.

Second, if the probability of missingness is dependent on the probability of the observed values in the data set, it is missing at random (MAR). Under MAR, the probability of missingness of X_{ikg} is specified as

$$\phi_{ikg} = p\left(M_{ikg} = 1 \middle| X_{ikg}, \phi\right) = p\left(M_{ikg} = 1 \middle| X_{obs}, \phi\right) \, for \, \forall \, X_{miss}, \phi \qquad (16)$$

Third, if the probability of missingness is dependent on the probability distribution of the missing values in the data set, it is missing not at random (MNAR). Under MNAR, the probability of missingness of X_{ikg} is specified as

$$\phi_{ikg} = p\left(M_{ikg} = 1 \middle| X_{ikg}, \phi\right) = p\left(M_{ikg} = 1 \middle| X_{miss}, \phi\right) \, for \, \forall \, X_{obs}, \phi \qquad (17)$$

For example, suppose that a student is administered two math exams. However, the student's score on the second math exam is missing. If the reason the student's second math exam score is missing is that they were absent on the exam day, this is a case of MCAR, since the missingness occurred by accident. If instead the student does not need to take the second math exam because the first math exam score satisfied the cutoff score, the missing data is a case of MAR, because the reason for missingness is dependent on the score of the prior test (i.e., the pretest) score. In yet another scenario, if the teacher decides not to record a score lower than the cut score on the second math exam, and the student does not score above the cut score, that is a case of MNAR. Because the reason for the missingness is dependent on the variable of interest (i.e., the student's performance on the second math exam), MNAR is not ignorable missing data (Rubin, 1976).

Missing Data Techniques: How to Deal with Missing Data

According to Roth (1994), missing data would be a potential problem when the researcher conducts the data analysis because missingness causes the loss of information about the targeted population. If the entire data set has the missingness, the statistical power of the results would be decreased, and the parameter estimates would be biased. Thus, researchers have to consider how to deal with missing data for diverse statistical analysis, such as non-normal data analysis (Enders, 2001a), structural equation modeling (Cai et al., 2009; Lu et al., 2011), and multilevel analysis (Black et al., 2011; Grund et al., 2018).

The literature has proposed diverse missing data techniques, which is a way to handle the missing values in the data set, because the missing data have a significant impact on the results of a statistical analysis (Allison, 1987; Enders, 2011; Little & Rubin, 2002). There are three categories of missing data techniques: data deletion method, data imputation method, and data augmentation method.

First, researchers commonly use two data deletion methods: the listwise deletion method (i.e., the complete case method) and the pairwise deletion method (i.e., the available case method). Suppose there is a data set of five students on four exams as follows:

	English Exam 1	English Exam 2	Math Exam 1	Math Exam 2
Student 1	54	92	70	94
Student 2	67	56		52
Student 3	81	77		
Student 4	90	87	64	74
Student 5	76	53	82	91

where black boxes indicate the missing values of the test scores. Using the listwise deletion (LD) method, the researcher uses the data set with three students (Students 1, 4, and 5) because the scores for math exams 1 and 2 are missing for two students (Students 2 and 3). But using the pairwise deletion method, the researcher analyzes the data set with two variables (English exams 1 and 2) because the scores of math exams 1 and 2 have the missing data.

However, as mentioned in the literature, the researchers should be careful in using either of these data deletion methods regarding missing data mechanisms (Little & Rubin, 2002; McKnight et al., 2007; Schafer & Graham, 2002). Under MCAR, because the existence of missing data is completely random, the data set after deleting missing values is one random sample of the original data set. Thus, the results of statistical analysis from the data set after deleting missing values are unbiased. However, under MNAR, data deletion methods could cause biased results of statistical analysis, reduce the statistical power, and invalidate the conclusions of the study.

Second, the data imputation method is a way to replace missing data with imputed values. The main goal of the data imputation method is to find the most plausible value for the replacement of the missing data. There are two data imputation methods. The single imputation method substitutes the missing data with a single value. For the replacement value, the researcher can consider using the arithmetic mean or the median of the observed values, or the number zero or the value derived from the regression analysis based on the variables with fully observed values. However, the single imputation method can cause biased results for the statistical analysis as well. Using the zero-imputation method, the imputed data set could be zero-inflated, which could be substantially different from the original data set. To use the regression imputation method, if the original data set has many variables with complicated missing patterns together, it is difficult to calculate the imputed values from the regression.

The multiple imputation method is a way to substitute the missing data with multiple values obtained from the repeated imputation processes instead of from only a single process (Sinharay et al., 2001; van Buuren, 2011, 2018). From Tanner and Wong (1987), the imputation model is specified as

$$f(X_{miss}|X_{obs}) = \int f(X_{miss}|X_{obs}, \theta, \phi) f(\phi|X_{obs}, \theta) d\phi \tag{18}$$

where X_{obs} is a matrix of the observed values in the data set of the score X_{ikg}, X_{miss} is a matrix of the missing values in the data set of the score X_{ikg}, θ is a set of parameters for the observed data distribution, and ϕ is a set of parameters for the missing data distribution. As we can see, the imputation model consists of two components: $f(X_{miss}|X_{obs}, \theta, \phi)$, which explains how to generate the values to replace X_{miss}, and $f(\phi|X_{obs}, \theta)$, which shows how to generate the parameter values of X_{miss} given X_{obs}.

Peugh and Enders (2004) mentioned that under MCAR or MAR, the multiple imputation method yields unbiased and efficient parameter estimates. In addition, the multiple imputation method is more flexible in applying the statistical algorithm to construct the model for the imputation process under complex data conditions, such as a non-normal data set. However, McKnight et al. (2007) pointed out that the researcher could arbitrarily set the model used for the multiple imputation method.

If the researcher uses either the data deletion method or the data imputation method, the data set with missingness is converted into the complete data. Finkbeiner (1979) showed how to find the parameter estimates of the factor analysis model with missing data as follows. Suppose that K examinees are administered a test with I test items. Then, the scores of all examinees on the test as $X = (x_{ik})$ follow multivariate normal distribution where $i\ (= 1, \dots, I)$ indicating each item and $k\ (= 1, \dots, K)$ indicating each examinee. If the sample size is large enough, the sample covariance $\boldsymbol{S}$ approximates the population covariance matrix Σ, which is Matrix 1 shown in the previous section as

$$\boldsymbol{S} \rightarrow \Sigma = \begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_{I-1} \\ X_I \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & \dots & X_{I-1} & X_I \end{array} \\ \boxed{\begin{array}{ccccc} \sigma^2_{X_1} & \sigma_{X_1X_2} & \dots & \sigma_{X_1X_{I-1}} & \sigma_{X_1X_I} \\ \sigma_{X_2X_1} & \sigma^2_{X_2} & \dots & \sigma_{X_2X_{I-1}} & \sigma_{X_2X_I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{X_{I-1}X_1} & \sigma_{X_{I-1}X_2} & \dots & \sigma^2_{X_{I-1}} & \sigma_{X_{I-1}X_I} \\ \sigma_{X_IX_1} & \sigma_{X_IX_2} & \dots & \sigma_{X_IX_{I-1}} & \sigma^2_{X_I} \end{array}} \end{array}$$

where $i = 1, 2, \dots, I$ is the number of indicators, $\sigma^2_{X_i}$ is the observed score variance of the indicator i, and $\sigma_{X_iX_j}$ is the covariance of the observed scores of the indicators i and j for $i \neq j$. Then, if the model is congeneric (i.e., we relax both assumptions of the same factor loadings and the same error variances; (a) $\lambda_1 \neq \lambda_2 \neq \cdots \neq \lambda_I \neq \lambda$, (b) $\theta_{11} \neq \theta_{22} \neq \cdots \neq \theta_{II} \neq \theta$), the predicted covariance matrix Σ_θ is Matrix 2 (i.e., the sum of Matrix 3 and Matrix 4) which is shown in the previous section as

$$\Sigma_\theta = \Lambda\Psi\Lambda' + \Theta = \begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_{I-1} \\ X_I \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & .. & X_{I-1} & X_I \end{array} \\ \boxed{\begin{array}{ccccc} \lambda_1^2 & \lambda_1\lambda_2 & .. & \lambda_1\lambda_{I-1} & \lambda_1\lambda_I \\ \lambda_2\lambda_1 & \lambda_2^2 & .. & \lambda_2\lambda_{I-1} & \lambda_2\lambda_I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{I-1}\lambda_1 & \lambda_{I-1}\lambda_2 & .. & \lambda_{I-1}^2 & \lambda_{I-1}\lambda_I \\ \lambda_I\lambda_1 & \lambda_I\lambda_2 & .. & \lambda_I\lambda_{I-1} & \lambda_I^2 \end{array}} \end{array} + \begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_{I-1} \\ X_I \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & \dots & X_{I-1} & X_I \end{array} \\ \boxed{\begin{array}{ccccc} \theta_{11} & 0 & \dots & 0 & 0 \\ 0 & \theta_{11} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \theta_{I-1I-1} & 0 \\ 0 & 0 & \dots & 0 & \theta_{II} \end{array}} \end{array}$$

where Λ is a matrix of factor loadings with the elements λ_i; Ψ is a matrix of factor variance-covariance, which is an identity matrix because the variance of a single factor is fixed as 1 for simplicity; and Θ is a matrix of error variances with the elements θ_{ii} with i indicating the ith item ($i = 1, \dots, I$).

The primary goal of the factor analysis model is to find the parameter estimates, which minimizes the difference between $\boldsymbol{S}$ and Σ_θ. Thus, the goal of the factor analysis model is to test a hypothesis that the sample covariance matrix equals the predicted covariance matrix based on the model. In the matrix format, using the factor analysis, we test a null hypothesis as H_0: $\boldsymbol{S} = \Sigma_\theta$ where θ is a set of parameters, Σ is the population covariance matrix, $\boldsymbol{S}$ is the sample covariance matrix, and Σ_θ is the predicted covariance matrix.

To estimate the parameters of the factor analysis model, the maximum likelihood (ML) method is commonly used (Brown, 2015; Marcoulides & Schumacker, 2013). For the kth examinee, the density of the scores is

$$f(\boldsymbol{x}_k|\theta) = (2\pi)^{-I/2}|\Sigma_\theta|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{x}_k{\Sigma_\theta}^{-1}{\boldsymbol{x}_k}'\right] \tag{19}$$

where $\boldsymbol{x}_k = [x_{k1}, \dots, x_{kI}]$ indicates a row vector of scores of the kth examinee of the test with I items, θ is a set of parameters, and Σ_θ is the predicted covariance matrix based on the model. Then, the likelihood is

$$L(\theta|\boldsymbol{x}_k) = \prod_{k=1}^{K} f(\boldsymbol{x}_k|\theta) = (2\pi)^{-KI/2}|\Sigma_\theta|^{-K/2} \exp\left[-\frac{1}{2}\sum_{k=1}^{K}\boldsymbol{x}_k{\Sigma_\theta}^{-1}{\boldsymbol{x}_k}'\right] \tag{20}$$

For simplicity of computation, instead of likelihood, we consider the log-likelihood, shown as

$$LL_{\Sigma_\theta} = \ln L(\theta|\boldsymbol{x}_k) = -\frac{KI}{2}\ln(2\pi) - \frac{K}{2}\ln(|\Sigma_\theta|) - \frac{1}{2}\sum_{k=1}^{K}\boldsymbol{x}_k{\Sigma_\theta}^{-1}{\boldsymbol{x}_k}' \tag{21}$$

This is a log-likelihood based on Σ_θ. Similarly, the log-likelihood based on $\boldsymbol{S}$ is represented as

$$LL_{\boldsymbol{S}} = \ln L(\theta|\boldsymbol{x}_k) = -\frac{KI}{2}\ln(2\pi) - \frac{K}{2}\ln(|\boldsymbol{S}|) - \frac{1}{2}KI \quad (22)$$

Regarding the hypothesis tested by the factor analysis model (i.e., H_0: $\boldsymbol{S} = \Sigma_\theta$), ML estimation aims to derive a set of the most probable values of the parameters that minimize the difference between Σ_θ and $\boldsymbol{S}$. Then, the difference between two matrices could be represented by the difference between LL_{Σ_θ} and $LL_{\boldsymbol{S}}$, shown as

$$LL_{\Sigma_\theta} - LL_{\boldsymbol{S}} = -\frac{K}{2}\left[\ln(|\Sigma_\theta|) - \ln(|\boldsymbol{S}|) + trace\left[(\boldsymbol{S})\left(\Sigma_\theta^{-1}\right)\right] - I\right] \quad (23)$$

where $\sum_{k=1}^{K}\boldsymbol{x}_k{\Sigma_\theta}^{-1}\boldsymbol{x}_k{}' = K \times trace\left[(\boldsymbol{S})\left(\Sigma_\theta^{-1}\right)\right]$. Because $-\frac{K}{2}$ is a constant that does not influence $LL_{\Sigma_\theta} - LL_{\boldsymbol{S}}$, the difference between the observation and the model is represented by a fit function shown as

$$F_{ML} = \ln|\Sigma_\theta| - \ln|\boldsymbol{S}| + trace\left[(\boldsymbol{S})\left(\Sigma_\theta^{-1}\right)\right] - I \quad (24)$$

where I indicates the total number of items in the test. This is a discrepancy function for ML estimation of the factor analysis model (Brown, 2015).

The third category of missing data techniques is the data augmentation method, which is a way to estimate the parameters of interest by augmenting the information from observed data and an underlying probability model (McKnight et al., 2007). This method differs from the data deletion method and the data imputation method in that it handles the incomplete data not by deleting or replacing the missing data but by directly estimating the parameters of interest through statistical inference.

One of the commonly used data augmentation methods is full information maximum likelihood (FIML) estimation (Allison, 2003; Baraldi & Enders, 2010; Little & Rubin, 2002; Schafer & Graham, 2002). Different from the data deletion method or the data imputation method, the data augmentation method derives the parameter estimates from the incomplete data without deleting or replacing the missing data.

Suppose that K examinees are administered a test with I test items. We can define the scores of all examinees on the test as $X = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_K]$ where $\boldsymbol{x}_k = [x_{k1}, \dots, x_{kI}]$ is a row vector of scores of the kth examinee of the test with I items. The resulting matrix represents an independent and identically distributed sample of K examinees from multivariate normal distribution. For the kth examinee, the density of the observed scores is

$$f(\boldsymbol{x}_k|\theta, \phi) = f\left(\boldsymbol{x}_k^{obs}, \boldsymbol{x}_k^{miss}\middle|\theta, \phi\right) = f\left(\boldsymbol{x}_k^{obs}\middle|\theta\right) f\left(\boldsymbol{x}_k^{miss}\middle|\boldsymbol{x}_k^{obs}, \theta, \phi\right) \tag{25}$$

where $\boldsymbol{x}_k = [x_{k1}, \dots, x_{kI}]$ indicates a row vector of scores of the kth examinee of the test with I items, θ is a set of parameters for observed data $\boldsymbol{x}_k^{obs}$, and ϕ is a set of parameters for missing data $\boldsymbol{x}_k^{miss}$. Thus, a model of the entire data $f(\boldsymbol{x}_k|\theta, \phi)$ is decomposed by $f\left(\boldsymbol{x}_k^{obs}\middle|\theta\right)$, which is a probability distribution of the observed data given the parameter θ, and $f\left(\boldsymbol{x}_k^{miss}\middle|\boldsymbol{x}_k^{obs}, \theta, \phi\right)$, which is a probability distribution of the missing data given the observed data and the parameter ϕ. Then, the likelihood is

$$L(\theta, \phi|\boldsymbol{x}_k) = \prod_{k=1}^{K} f(\boldsymbol{x}_k|\theta, \phi) = \prod_{k=1}^{K} f\left(\boldsymbol{x}_k^{obs}\middle|\theta\right) \prod_{k=1}^{K} f\left(\boldsymbol{x}_k^{miss}\middle|\boldsymbol{x}_k^{obs}, \theta, \phi\right) \tag{26}$$

The likelihood has two parts. $\prod_{k=1}^{K} f\left(\boldsymbol{x}_k^{obs}\middle|\theta\right)$ represents the likelihood of the observed data denoted as $\ln L\left(\theta|\boldsymbol{x}_k^{obs}\right)$, and $\prod_{k=1}^{K} f\left(\boldsymbol{x}_k^{miss}\middle|\boldsymbol{x}_k^{obs}, \theta, \phi\right)$ indicates the likelihood of the missing data. Then, the log-likelihood is

$$LL = \ln L(\theta, \phi|\boldsymbol{x}_k) = \ln L\left(\theta|\boldsymbol{x}_k^{obs}\right) + \ln\left(\sum_{k=1}^{K} f\left(\boldsymbol{x}_k^{miss}\middle|\boldsymbol{x}_k^{obs}, \theta, \phi\right)\right) \tag{27}$$

Then, the discrepancy function under FIML is

$$F_{FIML} = \left[\ln\left|\Sigma_\theta^{\mathrm{obs}}\right| - \ln|\boldsymbol{S}^{obs}| + trace\left[(\boldsymbol{S}^{obs})\left(\left(\Sigma_\theta^{obs}\right)^{-1}\right)\right] - I^{obs}\right]$$
$$+ \left[\ln\left|\Sigma_\theta^{miss}\right| - \ln\left|\boldsymbol{S}^{miss}\right| + trace\left[\left(\boldsymbol{S}^{miss}\right)\left(\left(\Sigma_\theta^{miss}\right)^{-1}\right)\right] - I^{miss}\right] \tag{28}$$

where $\Sigma_\theta^{\mathrm{obs}}$ is the predicted covariance matrix, $\boldsymbol{S}^{obs}$ is the sample covariance matrix, and I^{obs} is the number of items for observed data, while for missing data, Σ_θ^{miss} is the predicted covariance matrix, $\boldsymbol{S}^{miss}$ is the sample covariance matrix, and I^{miss} is the number of items. However, we cannot derive the components of the missing data because we cannot observe them directly. Thus, the discrepancy function under FIML is simplified as

$$F_{FIML} = \left[\ln\left|\Sigma_\theta^{\mathrm{obs}}\right| - \ln|\boldsymbol{S}^{obs}| + trace\left[(\boldsymbol{S}^{obs})\left(\left(\Sigma_\theta^{obs}\right)^{-1}\right)\right] - I^{obs}\right] \tag{29}$$

which we estimate the parameters of the factor analysis model when using FIML method to handle the missing values in the data set.

Effects of Missing Data on Reliability Estimation

The results of the empirical data analysis determined that missing values in the data set are one of the most important factors for the reliability estimation. Raykov (2009a) analyzed the small sample (219 elderly people's data) with 26.48% missing values using a single-factor model with four indicators. By comparing two missing data techniques (i.e., the listwise deletion [LD] method and full information maximum likelihood [FIML] method), the highest value was the estimate of the coefficient omega using the FIML method (.96), while the lowest value was the estimate of the coefficient alpha using the LD method (.69). In addition, by analyzing 200 teacher candidates' data set with up to 20% missing values, Çokluk and Kayri (2011) examined the performance of the coefficient alpha estimates under a single-factor model with 10 indicators. The estimates of the coefficient alpha varied (from .78 to .91) depending on the five data imputation methods (e.g., mean imputation, median imputation). Thus, through the results of these empirical data analyses, it is obvious that the missing values in the data set had significant effects on the reliability estimation, and the conditions related to the missing values (e.g., missing data proportions, or how many missing values were included in the data set; and missing data techniques, or how to handle the missing values) were the meaningful factors for the reliability estimation.

To specify how the missing data conditions affect the reliability estimation, researchers have conducted data simulation studies as follows. For the coefficient alpha estimates, Downey and King (1998) manipulated missing item proportions from 10% to 70% and the proportions of missing person's responses from 5% to 35%. Under these two missing data techniques (i.e., item mean imputation and person mean imputation methods), the estimates of the coefficient alpha showed .02 to .05 differences with the coefficient alpha based on the complete (0% missing) data set.

McDonald et al. (2000) conducted a simulation study manipulating the sample sizes (100 and 200), the number of items (3 and 6 items), the population values of the coefficient alpha (.70 and .90), and the inter-item correlations (.28, .44, .60, and .75). To generate missing completely at random (MCAR) missing data, they considered three conditions of missing data proportions (0%, 20%, and 40%). They used five missing data techniques (LD, item mean imputation, person mean imputation, random number imputation, and multiple linear regression imputation). The results showed that because (a) the missing data proportions increased, (b) the number of sample sizes and items decreased, and (c) the population value of the reliability was low, the coefficient alpha estimates were more biased. The LD method showed the highest bias of the five missing data techniques.

From a series of simulation studies, Enders (2003, 2004) examined the impact of the missing values on the reliability estimation by manipulating missing data proportions (15% and 30%), missing data mechanisms (MCAR, missing at random [MAR], and missing not at random [MNAR]), the number of samples (100, 300, and 500 samples), the number of items (10 and 20), the number of item response categories (3-, 5-, and 7-point scales), and inter-item correlations (.30 and .56). By comparing the results of five missing data techniques (LD, pairwise deletion, item mean imputation, person mean imputation, and the FIML method), (a) the estimates of the coefficient alpha showed the largest bias when using the LD method, while (b) the bias of the estimates of single-level coefficient alpha using the FIML method was the lowest under the MCAR condition, and (c) both the FIML and person mean imputation methods showed similar superiority in the reliability estimation under the MNAR condition.

By manipulating the conditions of the missing data proportions (5%, 10%, 20%, and 30%) and the sample sizes (200 and 3,000 samples) under MAR, Cuesta Izquierdo and Fonseca Pedrero (2014) established that (a) as the missing data proportions were increased, the bias of the coefficient alpha estimates increased, and (b) the usage of the LD method showed the highest bias of the different data imputation methods, including multiple linear regression imputation, imputation using expectation-maximization (EM) algorithm, and multiple imputation using fully conditional specifications.

For the coefficient omega estimates, Raykov (2009b) generated the 300 samples with a single-factor model with five items and estimated the coefficient omega estimates. After removing the last (i.e., the fifth) item, they calculated coefficient omega estimates using the FIML method. The difference between the results of the two cases was -.03, which indicates that the FIML method is robust to estimate reliability when the missing values are caused by item removal (i.e., change in the number of items or change of the test length).

To sum up, the results of the empirical and simulation studies showed that the missing data mechanisms (MCAR, MAR, MNAR), the missing data proportions (up to 40%), the missing data techniques (LD method, FIML method, and data imputation methods such as person mean imputation, item mean imputation, or multiple linear regression imputation), the population values of the reliability estimates, the sample sizes (from 100 to 3,000 samples), and the number of items in the test (3 to 20 items, i.e., the test length) were the significant factors for the reliability estimation with missing values in the data set. These conditions are important for the reliability estimation under the complete (0% missing) data set as well (Deng & Chan, 2017, Duun et al., 2014; Edward et al., 2021; Zinbarg, et al., 2005).

However, the aforementioned studies had some limitations: (a) the single-level reliability measures were examined, even though the multilevel data structure is prevalent in the field of education (Bryk & Raudenbush, 1987; Muthén, 1994, 1997; Raudenbush & Bryk, 2002), and (b) either the coefficient alpha or the coefficient omega was investigated, thus, the performance of both coefficient indices were not compared.

Researchers of multilevel reliability treated the missing data using one of the missing data techniques, such as the FIML method (Jak & Jorgensen, 2017; Jeon et al., 2009; Sideridis et al., 2018) or the LD method, which was not recommended because of its poor performance in the reliability estimation (Geldhof et al., 2014; Lai, 2021). Some of the multilevel reliability studies' authors did not investigate the effects of the conditions related to the missing data (Bonito et al., 2012).

CHAPTER 3

EMPIRICAL DATA ANALYSIS

For empirical data analysis, we analyzed the data set of the 2018 Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA 2018). The program was designed to measure the knowledge and skills of 15-year-old-students in reading, mathematics, and science. The data included the responses of 4,838 students from 164 schools, identified as clusters, in the United States. The cluster sizes, which indicate the number of students from each school, ranged from 1–39. The data were collected by the computer-based test administrations.

The variables of interest in this study included asking for general information about the students, their family, and their school. Also included were questions related to language learning in school, their view on reading, their perspective on their life vs. school schedule, and learning time. We chose a set of six items to measure how often the students thought about the other students in the same school (ST038, “During the past 12 months, how often have you had the following experiences in school?”). From the questionnaire about their familiarity with information and communications technologies (ICT), we selected another set of six items about the students’ experience with digital media and digital devices (IC013, “Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?”). Table 1 shows the 12 items of two sets used for the empirical data analysis in this study.

As shown in Table 2, the selected items included four response categories. For the six items of ST038, response 1 indicated “never or almost never,” and response 4 indicated “once a week or more.” For the six items of IC013, response 1 indicated “strongly disagree” and response 4 indicated “strongly agree.” The missing values of 12 items included the cases of system missing (i.e., the students’ answers were missing because of response time limit), not applicable (i.e., the students’ answers were missing because of the survey design), and no response (i.e., the students reached the item but did not answer). The missing data proportions of the six items of ST038 ranged from 5.4%–5.8%, and the six items of IC013 had missing values ranging from 8.8%–9.6%.

Among the six items of ST038, the fourth item (“Other students took away or destroyed things that belonged to me”) showed the lowest response mean value (M=1.20 with S.D.=.60), while the second item (“Other students made fun of me”) showed the highest response mean value (M=1.70 with S.D.=.90). However, more than half of the students answered the response 1 category (“never or almost never”), which indicated that they rarely had negative experiences with other students in the previous 12 months.

For IC013, the last item (“I like using digital devices”) had the highest response mean (M=3.20 with S.D.=.70), whereas the first item (“I forget about time when I’m using digital devices”) had the lowest response mean (M=2.60 with S.D.=.80). Because more than half of the students answered ‘agree’ or ‘strongly agree’, the results represented that the students were familiar with the digital media and digital device with positive thoughts of its usefulness.

Estimation

To determine the reliability indices (i.e., the single-level coefficient alpha α, the within-group level coefficient alpha α_w, the between-group level coefficient alpha α_B, the single-level coefficient omega ω, the within-group level coefficient omega ω_w, and the between-group coefficient omega ω_B), we hypothesized that the estimated results were either similar or different depending on how many missing values were included in each item (i.e., missing data proportions) and how to handle the missing values (i.e., missing data techniques).

We fitted the single-level confirmatory factor analysis (CFA) model to estimate the single-level reliabilities (α, ω), and the two-level CFA model to estimate the level-specific reliability indices (α_w, α_B, ω_w, ω_B) by treating the schools as the between-group level and the students as the within-group level in the data set. According to Geldhof et al. (2014), the CFA model to estimate the single-level and the level-specific coefficient alpha measures were fully saturated, which indicated that the model was simply identified with no degrees of freedom; thus, the model-fit of these two models was perfect (i.e., comparative fit index [CFI]=1, Tucker–Lewis index [TLI]=1, and root-mean-square error of approximation [RMSEA]=0). However, the CFA models for the coefficient omega were under-identified (e.g., there were certain degrees of freedom); thus, we set a criterion for the "good" model fit as per Geldhof et al. (2014): CFI > .90, TLI > .90, and RMSEA < .08.

Results

As shown in Table 3, for ST038, the CFA models to estimate the single-level and level-specific coefficient omega did not show good model-fit results, except for CFI of the single-level CFA model for ω using the LD method (CFI=.906) and FIML method (CFI=.908). However, for IC013, the models showed good model-fit results to estimate the coefficient omega, except the RMSEA of the CFA models using LD method (RMSEA=.091) and FIML method (RMSEA=.089).

From Table 3, for ST038, the estimates of the reliabilities were increased from using the LD method (α=.859, α_w=.856, α_B=.940, ω=.847, ω_w=.849, ω_B=.754) to using the FIML method (α=.860, α_w=.858, α_B=.944, ω=.848, ω_w=.850, ω_B=.833), while for IC013 the reliability estimates were decreased from LD (α=.825, α_w=.821, α_B=.928, ω=.808, ω_w=.805, ω_B=.947) to FIML (α=.823, α_w=.819, α_B=.924, ω=.806, ω_w=.803, ω_B=.944). The larger difference in the reliability estimates was observed in the estimates of ω_B of ST038 (difference of ω_B=.079) between the two missing data techniques.

By the definition mentioned by Raudenbush and Bryk (2002), the intra-class correlation (ICC) represents the ratio of the variability at the between-group level over the total (both within- and between-group level) variability, thus, the larger ICC indicates that the usage of the multilevel model technique is relatively appropriate for the data analysis. In addition, as mentioned by Little and Rubin (2002), when we use the LD method, the results of the statistical analysis could be affected because it removes the missing data from the entire samples. Thus, the combinations of the small ICC values and types of missing data techniques had an impact on the reliability estimation. Because ICCs were very close to zero for ST038 (ICC=.003 using LD, ICC=.005 using FIML), the usage of the two-level MCFA was not proper. Thus, the results showed that there was an impact of the magnitudes of ICC values and the missing data techniques on the reliability estimation.

However, except for ω_B of ST038, the results of other conditions were not substantially different because the data conditions of ST038 and IC013 were not in common. From the relevant literature (Hox, 2010; Kim & Cao, 2015; Koch et al., 2014; Snijders & Bosker, 2012), the ICC values ranging from .10–.20 have been commonly observed in most educational studies, and .30 is the maximum ICC value in psychological studies. Even though the ICC of IC013 was larger (ICC=.029 using LD and FIML) than the ICC of ST038 (ICC=.003 using LD, ICC=.005 using FIML), they were smaller than .10. In addition to ICCs, as reviewed by Li and Lomax (2007), the missing data proportions, which ranged from 0%–30%, were commonly examined in the latent variable analysis in the educational and psychological studies. However, as shown in Table 3, the items of ST038 and IC013 had below 10% missingness. Thus, the effects of ICCs' missing data proportions and missing data techniques might be trivial in the reliability estimation of both sets.

CHAPTER 4

SIMULATION STUDIES

Data Generation

Data Sets Without Missing Values

Our study generated the data based on an MCFA model with one factor and six indicators, as shown in Figure 2. According to Raykov (2012), the MCFA model in this study showed the test has unidimensionality. Following the simulation conditions that Geldhof et al. (2014) examined, the model did not have the covariates, the indicators were the continuous variables, and the intercepts were fixed at zero for simplicity. Additionally, we assumed the model was the congeneric model (i.e., the factor loadings and the error variances could differ) at within- and between-group levels, respectively. The congeneric model is the most flexible case regarding the assumptions for reliability estimation (Brown, 2015; Cho, 2016) and the most general case for factor analysis techniques to estimate the latent structure of the measurement tool (Jöreskog, 1971; Raykov & Hancock, 2005).

The two panels in Figure 2 indicate the MCFA model was under a two-level data structure: the lower panel was the within-group level model, and the upper panel was the between-group level model. In the within-group level model, the factor was shown as the circle in the lower plate with the factor score $\eta_w \sim N(0,1)$ because it was an unobservable (i.e., latent) variable. To scale the factor for model identification, we fixed the variance of the factor as 1 rather than fixing one of the factor loadings as 1 (Raykov & Hancock, 2005; Raykov & Marcoulides, 2006). The relationship between the factor and the six indicators X_i $(i = 1, \ldots, 6)$

was represented as the solid line with the factor loadings λ_{wi} and the corresponding errors of each indicator $\varepsilon_{wi} \sim N(0, \theta_{wii})$. Each error was shown in the circle to represent each indicator's unique features (i.e., the part unexplained by the factor). The error variances were non-zero value θ_{ii} because they were allowed to include the difference among the examinees and to deal with the assumptions for the congeneric test under the factor analysis framework (Li et al., 1996).

We dealt with the presence of the missing values; thus, M_{X_i} in the square indicated whether the corresponding indicator X_i was observed. We represented the probability density of missing data $p\left(M_{X_i}\right)$ with the dashed line between X_i and M_i determined by the manipulated conditions to generate the missing data.

At the between-group level, a single factor was represented as the circle of the factor score $\eta_B \sim N(0,1)$ related to the six indicators with factor loadings λ_{Bi}. Different than the indicators at the within-group level, the between-group level indicators X_{Bi} were shown as the latent variables because X_{Bi} indicated the shared but unobservable features of the within-group level units belonging to the same between-group level unit. Their corresponding errors were $\varepsilon_{Bi} \sim N(0, \theta_{Bii})$ as seen in the circles linked with the arrows in Figure 2.

Because the generated data had a two-level structure, we manipulated three ICC conditions, as shown in Table 4. Raudenbush and Bryk (2002) defined ICC as the ratio of the between-group variance over the total variance. From Muthén (1994), the ICC for an MCFA with a two-level data structure is as follows:

$$ICC = \frac{(\sum_{i=1}^{I} \lambda_{Bi})^2 + \sum_{i=1}^{I} \theta_{Bii}}{(\sum_{i=1}^{I} \lambda_{Bi})^2 + \sum_{i=1}^{I} \theta_{Bii} + (\sum_{i=1}^{I} \lambda_{wi})^2 + \sum_{i=1}^{I} \theta_{wii}} \tag{30}$$

where $i = 1, 2, \ldots, I$ is each of the test items, λ_{Bi} and λ_{wi} are the factor loadings, and θ_{Bii} and θ_{wii} are the error variances at each of the between- and within-group level models. If the ICC is large, then this indicates the component of the between-group level is large enough to explain the total variability. According to the literature, the ICC value ranges from .10 to .20 in the field of education, and the maximum value is .30 in psychological studies in general (Hox, 2010; Kim & Cao, 2015; Koch et al., 2014; Snijders & Bosker, 2012). Thus, we set three ICC conditions in this study: .050 as low, .111 as medium, and .296 as large.

We manipulated these three ICCs by the population values of factor loadings and error variances through the following steps. The previous studies treated the factor loadings from .60 to .90 as high, while from .20 to .50 as low (Green & Yang, 2009a); Hancock & An, 2020). First, we chose one of the conditions that Geldhof et al. (2014) examined as the combination of factor loadings and error variances at the within-group level as [($\lambda_{w1} = \lambda_{w2} =$ .599, $\lambda_{w3} = \lambda_{w4} =$ .699, $\lambda_{w5} = \lambda_{w6} =$ .799), ($\theta_{w11} = \theta_{w22} =$ .451, $\theta_{w33} = \theta_{w44} =$ .501, $\theta_{w55} = \theta_{w66} =$ .551)] and at the between-group level as [($\lambda_{B1} = \lambda_{B2} =$ .137, $\lambda_{B3} = \lambda_{B4} =$ .160, $\lambda_{B5} = \lambda_{B6} =$ .183), ($\theta_{B11} = \theta_{B22} =$ .019, $\theta_{B33} = \theta_{B44} =$ .027, $\theta_{B55} = \theta_{B66} =$ .034)]. The population value of this condition's ICC was .050. Second, we specified the combinations of factor loadings and error variances by fixing the population values as the same at the between-group level but different at the within-group level, depending on the three ICCs. Last, we set the rest of the population values as the same at the within-group level and different at the between-group level, depending on the ICCs. The population values of these factor loadings and error variances also determined the population values of the reliability measures. For the purposes of this study, we used these population reliability values (as shown in the Reliability column of Table 4) to evaluate the accuracy of the estimation. From the previous studies, because the reliability is an analogue of the correlation coefficient among the observed scores of the test items (Kelly, 1942), the values of the reliability from .20 to .40 considered as low, while from .70 to .90 considered as high (Edward et al., 2021; Bonito et al., 2012; DeVellis, 1991).

In addition to ICCs, we manipulated the sample sizes for data generation. Because of the two-level data structure, we considered the number of clusters and the cluster sizes. The number of clusters indicated how many between-group level units were collected in the data set. The cluster sizes represented how many within-group level units belonged to each cluster. For example, when we collected data from 50 schools and each school had 100 students, the number of clusters was 50, and the cluster size was 100. In our study, there were four conditions for sample sizes regarding the combinations of the number of clusters and cluster sizes (formulated as [(Number of Clusters, Cluster Sizes) = Total Sample Sizes]): [(50, 15) = 750], [(50, 30) = 1,500], [(100, 15) = 1,500], and [(100, 30) = 3,000]. Researchers commonly examined these sample size conditions in educational studies for multilevel modeling analysis (Boomsma, 1987; Duhachek & Iacobucci, 2004; Hsu et al., 2017; Loehlin, 2004; Padilla & Divers, 2013, 2016; Ryu & West, 2009).

We used the Mplus program, version 8 (Muthén & Muthén, 1998–2017), to generate the complete data sets using the aforementioned ICC and sample size conditions.

Data Sets With Missing Values

Our study generated data sets with missing values according to two features: missing data mechanisms and missing data proportions. These features were significant to the reliability estimation (Cokluk & Kayri, 2011; Estabrook & Neale, 2013; Enders, 2004; Leite & Beretvas, 2010). Similar to Enders's (2004) simulation design, we separated the six indicators into two groups: Group 1 (i.e., X_1 to X_3), which were the completely observed variables, and Group 2 (i.e., X_4 to X_6), which were the variables with missing values. We used these groups to manipulate the missing data mechanisms as follows.

We examined three types of missing data mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). For MCAR, we used Twala's (2009) missing data generation process. First, we created three binary variables b_4 to b_6 corresponding to the indicators X_4 to X_6 in Group 2 based on the Bernoulli distribution with the probability p (i.e., the missing data proportion). The value of each binary variable was either one or zero. If the value of the b variable was 1, then the value of the corresponding indicator was deleted; if the value of the binary variable was 0, then we treated the value of the corresponding indicator as the observed value.

To generate MAR and MNAR missing data, we used Garciarena and Santana's (2017) procedure. For MAR, we matched three pairs of the six indicators selected from Group 1 and Group 2 as (X_1, X_4), (X_2, X_5), and (X_3, X_6). We then sorted the Group 1 indicator values into ascending order. We omitted Group 2 indicator values when the corresponding indicator for Group 1 was ranked as the smallest p percentage. For MNAR, we sorted the Group 2 indicator values into ascending order. Next, we deleted Group 2 indicator values if the values were ranked in the smallest p percentage.

In addition to missing data mechanisms, we manipulated three missing data proportion conditions. The first condition was 0% as the complete data. The second and third conditions were 15% and 30%, which we used as p (as noted earlier). Because our study set the three indicators X_1 to X_3 as the completely observed variables and the other three indicators were X_4 to X_6 with missing values, the missing data proportions of the entire data set were 0%, 7.5%, and

15%. As found in the most educational studies, the missing proportions were below 30% in common (Cheema, 2014; Enders, 2003; Enders et al., 2006; Peugh & Enders, 2004).

Figures 3, 4, and 5 showed the procedures for generating the missing values under the three missing data mechanisms with 30% missing data proportions. In each figure, the left side shows the raw data of all six indicators X_1 to X_6 with 20 observations. The shaded area of each figure's right side shows the missing values of the indicators X_4 to X_6.

For MCAR in Figure 3, because we set the missing proportions at 30%, we created three binary variables, b_4 to b_6, set the six values as 1, and set the other 14 values as zero. Then, we deleted the six values of the indicators X_4 to X_6 depending on whether the value of the corresponding b variable was 1. Figures 4 and 5 showed the missing values of the indicator X_4 were generated under the MAR and MNAR mechanisms. As shown in Figure 4, for MAR, after sorting the values of X_1 into ascending order, we removed the six values of X_4 when the corresponding values of X_1 were the smallest six values. Figure 5 shows MNAR, in which we deleted the six values of X_4 when their values ranged in the smallest 30%.

Our study used the missMethods package, version 0.2.0 (Rockel, 2020; Santos et al., 2019), in R, version 4.1.2 (R Core Team, 2021), to generate the missing data sets in terms of the missing data mechanisms and missing data proportions.

As shown in Tables 5, 6, and 7, we examined 216 conditions for three missing data mechanisms (MCAR, MAR, and MNAR), four sample sizes (750, 1,500 with 100 groups, 1,500 with 50 groups, and 3,000 samples), six ICCs (.050, .111, and .296 with the same between-group level but different within-group level parameters and .050, .111, and .296 with the same within-group level but different between-group level parameters), and three missing data proportions (0%, 15%, and 30%). For each condition, we conducted 1,000 replications.

Data Analysis

To estimate the six reliability measures (i.e., single-level coefficient alpha α; level-specific coefficient alpha, including within-group level coefficient alpha α_W and between-group level coefficient alpha α_B; single-level coefficient omega ω; level-specific coefficient omega, including within-group level coefficient omega ω_W and between-group level coefficient omega ω_B), we used two missing data techniques to handle the generated data sets with missing values. Among three types of missing data techniques, the data deletion method has been a common choice in both the education and social sciences fields (Baraldi & Enders, 2010; Dong & Peng, 2013; Peugh & Enders, 2004; Rousseau et al., 2012). However, the data deletion method provides biased results under MAR conditions, because the data set cannot be sufficiently representative of the whole population after deleting the missing values (Schafer & Graham, 2002). More recently, researchers have preferred the data augmentation method to handle MAR missing values, which produces more accurate results in the statistical analysis, such as the latent variable model (Allison, 2003; Cham et al., 2017; Enders & Bandalos, 2001; Li & Lomax, 2017), longitudinal data analysis (Ibrahim & Molenberghs, 2009; Lu et al., 2011), or multilevel models (Goldstein, et al., 2014; Grund et al., 2019; Lüdtke et al., 2017). Thus, we used two missing data techniques—the listwise deletion (LD) method and the full information maximum likelihood (FIML) method—to compare their performance when handling the missing values in the reliability estimation.

In Mplus, we used the LD method by setting "LISTWISE=ON" under the DATA command. We set the FIML method to "MISSING ARE ALL (999)" in the VARIABLES command (see the Appendix for the Mplus codes used in our data analysis). These were the modified versions of the Mplus codes shown in Geldhof et al. (2014)—a one-factor MCFA model with four indicators under two levels—extended for our model with six indicators.

To evaluate the reliability estimation regarding the simulated conditions, we used two criteria. To evaluate the accuracy of the estimation, the first criterion was the percentage bias calculated as follows:

$$Percent\ Bias\ (\%) = \left[\frac{1}{1000}\sum_{r=1}^{1000}\frac{\left(\hat{\rho}_{XT_r} - \rho_{XT_r}\right)}{\rho_{XT_r}}\right] \times 100 \tag{31}$$

where $\hat{\rho}_{XT_r}$ was the estimated reliability measure in the rth replication, and ρ_{XT_r} was the population value under each simulated condition. Per Kaplan (1988) and Muthén et al. (1987), a magnitude of percentage bias below 15% could be considered an ignorable inaccuracy in the statistical estimation. Thus, we marked the condition showing bias above 15%, which indicated the condition would have a meaningful negative impact on the reliability estimation.

To evaluate the stability of the estimation, we calculated the second criterion, the convergence rate, as follows:

$$Convergence\ Rate\ (\%) = \frac{(Number\ of\ Converged\ Models)}{1000} \times 100 \tag{32}$$

Following Geldhof et al. (2014), we defined the model as converged (a) when the model satisfied the good model-fit values (i.e., comparative fit index [CFI] > .90, Tucker–Lewis index [TLI] > .90, and root-mean-square error of approximation [RMSEA] < .08) and (b) when there were no convergence error messages from the Mplus outputs. The larger value of the convergence rate indicated more stable results of the reliability estimation. Because the convergence rates under 50% meant half the replicated simulations failed to be converged, we marked the condition showing below 50% as convergence rates.

Results

The convergence rates and the percentage biases of the six reliability indices (i.e., single-level coefficient alpha α; level-specific coefficient alphas—that is, within-group level coefficient alpha α_W and between-group level coefficient alpha α_B; single-level coefficient omega ω; and level-specific coefficient omegas—that is, within-group level coefficient omega ω_W and between-group level coefficient omega ω_B) are shown in Tables 8–31. In each table, # represents the ID number of the simulated conditions listed in Tables 5–7. And MDM indicates three missing data mechanisms (i.e., MCAR, MAR, and MNAR), SS indicates four sample sizes (i.e., 750 samples, 1,500 samples with 100 clusters, 1,500 samples with 50 clusters, and 3,000 samples), ICC indicates three ICC values of either the same between-group parameters or the same within-group parameters, and MDP indicates three missing data proportions: 0%, 15%, and 30%. We reported the results of the convergence rates as the percentage % in the CR column and the percentage bias as the percentage % in the PB column under each condition.

In addition to tables, Figures 6, 9, 12, 15, 18, and 21 show the convergence rates of the model to estimate the single-level and the level-specific (i.e., the within-group and between-group level) coefficient omega, ω, ω_W, and ω_B. We did not report the convergence rates of the coefficient alpha, α, α_W, and α_B in the figures; the convergence rates were 100% under all conditions. As pointed out by Geldhof et al. (2014), the single-level CFA model and the two-level MCFA model to estimate the coefficient alpha were set as the fully saturated model, which was just identified with no degrees of freedom (Kline, 2016). Thus, we displayed the convergence rates of the models estimating the coefficient omega measures. To find the severe nonconvergence issue, each figure has dashed lines indicating 50% convergence rates.

The rest of the figures (Figures 7, 8, 10, 11, 13, 14, 16, 17, 19, and 20) show the percentage bias of the coefficient alpha and the coefficient omega. Each figure has solid lines indicating 0% bias to discriminate either the overestimated or the underestimated values of the reliability indices, and dashed lines indicating 15% bias to mark the severe lack of accuracy in the reliability estimation.

Results of MCAR Missing Values: LD Method

Figures 6–8 and Tables 8–11 report the convergence rates and the percentage biases of the coefficient alpha and the coefficient omega when analyzing the data with MCAR missing values with the listwise deletion (LD) method. In terms of the stability of the reliability estimation, all simulated conditions showed convergence rates above 50%. The lowest convergence rate (52.9%) was reported in the condition of the 750 samples with 30% missing values and the large ICC (.296) with the same between-group level parameters (i.e., Condition #9). This low convergence rate indicated that the researcher should be careful about the possibility of a nonconvergence issue if one of three conditions is present: (a) the sample is small, such as 750 samples with 50 clusters and 15 cluster sizes; (b) the missing data proportion is larger than 30%; or (c) the ICC is larger than .296 with the same (or similar) between-group parameters, that is, the population values of the within-group and between-group level reliabilities are different.

For the percentage bias, first, as the missing data proportions increased from 0% to 30%, the absolute values of the percentage bias increased. For example, under the condition of 750 samples with the small (.050) ICC values fixing the same between-group level parameters (i.e., Conditions #1–#3), the magnitudes of percentage bias of α_B increased from 1.5% to 4.7%, and those values of ω_B increased from 3.5% to 3.6%. Thus, when increasing the missing data proportions, the magnitudes of the inaccuracy of the reliability estimations increased.

Second, when the sample sizes increased from 750 to 3,000, the percentage bias of the single-level and the within-group level reliability estimates was increased. However, the percentage bias of the between-group level reliabilities, α_B and $\omega_{B,}$ decreased. For example, comparing the conditions of 750 samples (i.e., 50 clusters with 15 cluster sizes) and 1,500 samples (i.e., 50 clusters with 30 cluster sizes) with 30% missing values and the small ICC (.050) fixing the same between-group level parameters, that is, Condition #3 versus #12, the percentage bias decreased from 4.7% to 2.5% for α_B, and the percentage bias of ω_B decreased from 3.6% to 3.1%. In addition, when comparing the conditions of 750 samples and another 1,500 samples (i.e., 100 clusters with 15 cluster sizes) with 30% missing values, the small ICC (.050) fixing the same between-group parameters, that is, Condition #3 versus Condition #21, the absolute values of the percentage bias decreased from 4.7% to 2.5% for α_B. Thus, the results showed that as the sample sizes increased, the percentage bias of the reliabilities decreased.

Third, depending on the magnitudes of ICCs, the percentage bias of the reliability estimates showed the patterns as follow. When fixing the same between-group level parameters, as the ICC values increased from .050 to .296, the absolute values of the percentage bias of the single-level and the within-group level reliability estimates increased, but those absolute values of the between-group level reliabilities decreased. For example, in the condition of 750 samples with 30% missing values (i.e., among Conditions #3, #6, #9), the magnitudes of the percentage bias increased from 1.4% to 2.9% for α, 1.4% to 5.7% for α_w, 0.1% to 0% for ω, and 0% to 3.6% for ω_w. However, the percentage bias decreased from 4.7% to 1.5% for α_B and 3.6% to 0.2% for ω_B. When fixing the same within-group level parameters, as the ICC values increased

(.050 to .296), the magnitudes of the percentage bias increased for the single-level and the within-group level reliabilities. For example, under the condition of 750 samples with 30% missing values (i.e., among Conditions #39, #42, #45), the absolute values of the percentage bias increased from 1.4% to 2.1% for α and 1.4% to 1.8% for α_w. The values of ω and ω_w did not change significantly. Thus, when ICC was large, the between-group level reliabilities provided more accurate results, and when ICC was small, the single-level and the within-group level reliabilities showed smaller percentage bias.

Regarding the cutoff criterion for the percentage bias, there was only one result, that is, the percentage bias of α_B showed more than 15%, which indicated the serious problem in the accuracy of the reliability estimation under the condition of 750 samples with 30% missing values and the medium ICC (.111) fixing the same within-group level parameters (i.e., Condition #42). In addition, under the same condition, the percentage bias of ω_B (12.1%) was the lowest value among all simulated conditions for MCAR using the LD method. The results represented the factors that negatively affected the accuracy of the reliability estimation: the smaller the number of samples (750), the larger (30%) the missing data proportions, and the smaller the population values of the within-group reliabilities (α_w = .372, ω_w = .376 as shown in Table 4) to fix the same within-group level parameters for setting the medium (.111) ICC value.

Under most of the simulated conditions, the absolute values of the percentage bias of the coefficient alpha were larger than those of the percentage bias of the coefficient omega, and the percentage bias of the coefficient alpha were positive values, whereas the percentage biases of the coefficient omega were negative values. This result indicated that when the data had MCAR missing values and we applied the LD method, the estimates of the single-level and the level-specific coefficient alpha, α, α_w, and $\alpha_{B,}$ tended to be overestimated, whereas the estimates of the coefficient omega, ω, ω_W, and $\omega_{B,}$ were underestimated by the single-level CFA model and the two-level MCFA model.

Results of MCAR Missing Values: FIML Method

The results of the convergence rates and percentage biases of reliability estimates under MCAR using the full information maximum likelihood (FIML) method are shown in Figures 9–11 and Tables 12–15. All conditions showed the over 50% convergence rates when estimating the single-level and level-specific coefficient omega. We observed the lowest convergence rate (74.8%) in the condition of 750 samples with 30% missing values and the large ICC (.296) with the same between-group level parameters to estimate the within- and between-group level coefficient omega. Even though the same condition (Condition #9) showed the lowest convergence rate when using the LD method, the FIML method showed larger convergence rates than the LD method did. The results indicated that the use of the FIML method to handle the missing values would produce more stable estimations, with fewer convergence issues, than would use of the LD method.

As shown in Tables 12–15, the percentage biases of all six reliability estimates were not above 15% under all conditions. Regarding the missing data proportions, from 0% to 30%, the percentage biases of the coefficient alpha and the coefficient omega increased. For example, under the condition of 750 samples with the small (.050) ICC values fixing the same between-group level parameters (i.e., among Conditions #1–#3), the magnitudes of the percentage biases increased from 1.5% to 2.4% for α_B and 3.5% to 3.9% for ω_B. The results indicated that the missing data proportions had a negative impact on the accuracy of the reliability estimates.

In addition to the missing data proportions, as the sample sizes increased from 750 to 3,000, the percentage biases of the between-group level reliability estimates decreased, but the single-level and the within-group level reliability estimates did not change significantly. For example, under the conditions of the small (.050) ICC with the same between-group level parameters and 30% missing values, as the sample sizes increased from 750 (50 clusters and 15 cluster sizes) to 1,500 (100 clusters and 15 cluster sizes)—that is, Conditions #3 versus #12—the absolute values of the percentage bias of α_B decreased from 2.4% to 0%, and those of ω_B decreased from 3.9% to 1.2%. Under the conditions of 750 samples to another 1,500 (50 clusters and 30 cluster sizes)—that is, Conditions #3 versus #21—the magnitudes of the percentage biases decreased from 2.4% to 0% for α_B and 3.9% to 1.0% for ω_B. Thus, the small sample sizes reduced the accuracy of estimating the reliabilities, especially for α_B and ω_B.

When the between-group parameters were fixed as the same to set the ICCs and the population values of the reliabilities, as the ICCs increased the percentage biases of the single-level α and ω and the within-group level reliability estimates α_w and ω_w increased, whereas that of the between-group level reliability estimates α_B and ω_B decreased. For example, under the condition of 750 samples with 30% missing values, as ICC increased from .050 to .296 (i.e., Conditions #3, #6, #9), the absolute values of the percentage bias increased from 1.5% to 3.3% for α, 1.5% to 5.0% for α_w, 0.1% to 0.3% for ω, and 0% to 0.8% for ω_w. However, the magnitude of the percentage bias of α_B decreased from 2.4% to 1.3%, and the magnitude of the percentage bias of ω_B also decreased from 3.9% to 0.4%. Thus, the increase of ICCs had a positive impact on the accuracy of the between-group level reliabilities but a negative impact on the single-level and the within-group level reliability estimates.

When the within-group parameters were fixed as the same, as ICC was increased from .050 to .296 the percentage biases of the single-level and the within-group level reliabilities increased, whereas the percentage biases of the between-group level reliabilities decreased. The patterns of the results were similar under the same between-group parameters.

Regarding the cutoff for the percentage bias (i.e., above 15% bias is a serious problem in the estimation accuracy) under the condition of 750 samples with 30% missing values and the same within-group level parameters with the medium (.111) ICC (i.e., Condition #42), the percentage biases were the largest: 6.3% for α_B and 6.6% for ω_B. It is because Condition #42 combined the factors with the negative impact on the accuracy of the between-group level reliabilities, which included the smaller (750) samples with the larger (30%) missing data proportions and the different population values of the within- and between-group level reliabilities indices (α_w = .372, α_B = .850 for the coefficient alpha, ω_w = .376, and ω_B = .852 for the coefficient omega) to set the medium (.111) ICC fixing the same within-group level parameters.

Under most of the simulated conditions, the magnitudes of the percentage bias of the coefficient alpha were larger than the coefficient omega showed. However, the percentage biases of the coefficient alpha were positive values, which indicated that they tended to be overestimated, whereas the percentage biases of the coefficient omega were negative values, indicating underestimated values.

The results utilizing the FIML method showed similar patterns to those of the LD method. However, the convergence rates were larger, whereas the magnitudes of percentage bias were smaller when using the FIML method. For example, under Condition #42, the percentage bias of α_B was 19.2% using the LD method but 6.3% using the FIML method. Under Condition #9, the convergence rate of a two-level MCFA model to estimate the coefficient omega was 52.9% using the LD method but 74.8% using the FIML method. Hence, use of the FIML method provided more accurate results in the reliability estimation than the LD method. However, with the exception of Condition #42, there were no serious problems in the convergence rates (i.e., below 50%) and the percentage bias (i.e., above 15%) under MCAR conditions. As Rubin (1976) mentioned, we can treat the missing data generated by the MCAR mechanism as ignorable missing data because the missing data are not generated systematically. Thus, the impacts of MCAR missingness in the entire data set are small; thus, using either the LD or the FIML method works well to handle the missing data in the reliability estimation.

Results of MAR Missing Values: LD Method

Figures 12–14 and Tables 16–19 show the convergence rates and the percentage biases of the coefficient alpha, and the coefficient omega estimated from the analysis of the data with MAR missingness using the LD method. In terms of the stability of the estimation, as we mentioned in the previous section, because the coefficient alpha was estimated from the fully saturated model, the convergence rates were 100% under all simulated conditions. For the coefficient omega, most of the conditions showed the over 50% convergence rates. However, under the condition of 750 samples with 30% missingness and the larger ICC (.296) fixing the same between-group level parameters (i.e., Condition #81), the two-level MCFA model to estimate the within- and between-group level coefficient omega, ω_W and ω_B showed the lowest and below 50% convergence rate (39.8%).

Also under the same condition, the model to estimate the single-level coefficient omega showed the lowest convergence rate (70.0%). The lowest convergence rate was caused by the effects of the smaller samples with large missing data proportions and the large ICC value with fixing the same between-group parameters, which showed the lowest convergence rates under the conditions of MCAR. Thus, the result indicated that regardless of the missing data mechanisms (i.e., the reasons causing missing data), either MCAR or MAR, the combinations of the sample sizes, the missing data proportions, and ICC values have a negative impact on the convergence issue in the reliability estimation. However, because the same conditions of MCAR using the LD method showed a larger convergence rate (52.9%) than MAR using the LD method (39.8%), MAR had more negative impact on the convergence rates in the reliability estimation.

Comparing the conditions of 750 samples and 1,500 samples with 100 clusters and 15 cluster sizes under the small (.050) ICC with fixing the same between-group level parameters having 30% missing values (i.e., Conditions #75 versus #84), the absolute values of the percentage bias decreased from 10.6% to 10.4% for α_w and from 20.7% to 20.3% for α_B, and the percentage bias of α was 10.6% in each of the sample size conditions (750 and 1,500). However, under the same conditions, the percentage bias increased from 13.0% to 13.1% for ω, from 12.7 to 12.8% for ω_w, and from 17.3% to 19.6% for ω_B.

Also, comparing Condition #75 with another condition of 1,500 samples with 50 clusters and 30 cluster sizes (i.e., Condition #93), the absolute values of the percentage bias decreased for the coefficient alpha, from 10.6% to 10.5% for α_w, and from 20.7% to 20.0% for α_B. However, ω_w (from 12.7% to 12.8%), and ω_B (from 17.3% to 18.8%) showed increases in the percentage bias. The results indicated that the conditions with the larger sample sizes showed the lower percentage bias in the coefficient alpha estimates, while the higher percentage bias was reported the coefficient omega estimates.

Under the conditions of the same between-group level parameters, as ICC increased from .050 to .296, the percentage biases of the single-level and the within-group level reliabilities increased, while the percentage biases of the between-group level reliability estimates decreased. Among the conditions of 750 samples with 30% missing values by increasing ICCs from .050 to .296 (i.e., Conditions #75, #78, #81), the percentage biases increased from 10.6% to 20.1% for α, from 10.6% to 21.5% for α_w, from 13.0% to 22.5% for ω, and from 12.7% to 20.8% for ω_w. However, the percentage biases of α_B (20.7% to 12.2%) and ω_B (17.3% to 7.4%) decreased.

Under the conditions of the same within-group level parameters, the percentage bias of the single-level reliability estimates showed the same pattern with the conditions of the same within-group level parameters as the increase of ICCs. However, the percentage biases of the within- and between-group level reliabilities were different. For example, under the conditions of 750 samples with 30% missing values (i.e., Conditions #111, #114, #117), when increasing ICCs from .050 to .296, the percentage biases increased from 10.6% to 11.8% for α, from 20.7% to 46.7% for α_B, from 13.0% to 13.3% for ω, and from 17.3% to 44.0% for ω_B. However, among three ICC conditions, the percentage biases of α_w (10.6% to 6.6%) and ω_w (12.7% to 8.6%) decreased as ICCs increased. Thus, the results indicated that the increase of the ICC had a positive impact on the accuracy of the between-group level reliabilities under the conditions fixing the same within- and between-group level parameters.

Regarding the cutoff criterion for the percentage bias (i.e., above 15%), under the conditions of the same between-group level parameters, for the single-level and the within-group level reliabilities, the coefficient alpha, α and α_w, showed the percentage bias over 15% when the missing data proportions were 30% and ICC was .296. The coefficient omega, ω and ω_w, showed over 15% bias when the missing data proportions were larger than 15% and ICC was larger than .111. Under the conditions of the same between-group level parameters, when the missing data proportions were 30% and ICC was small (.050), the between-group level reliability estimates showed over 15% bias. However, under the conditions of the same within-group level parameters, when the missing data proportions were over 15%, the reliabilities at the between-group level, α_B and ω_B, showed a percentage bias larger than 15%.

Comparing the types of the reliability measures, both coefficient alpha and omega showed the negative percentage bias, which indicated the underestimated results under the conditions of 15% or 30% missing data proportions. However, under the conditions of the complete data (i.e., 0% missingness), the coefficient alpha showed the positive percentage bias, which indicated the overestimated results. Because LD method deleted the missing values from the data set, the reduction of the original data caused the underestimation for the single-level and the level-specific (i.e., the within-group level and the between-group level) reliabilities.

As shown in Figures 6 and 10, the percentage bias of the coefficient alpha under MAR was larger than under MCAR even though we used the same missing data technique, the LD method. Also, for the percentage bias of the coefficient omega, MAR relayed more inaccurate results than MCAR did. As pointed by Rubin (1976), even though MCAR and MAR are ignorable missingness, MAR generates the missing data depending on the other variables, not the variables of the interest. Thus, when utilizing LD method, the reduction in the data set is random in the variables of the interest, but not random in the other variables. For this reason, the usage of LD method reported larger percentage bias in the conditions of MAR than MCAR.

Results of MAR Missing Values: FIML Method

Figures 15–17 and Tables 20–23 show the convergence rates and the percentage bias of the coefficient alpha, and the coefficient omega estimated from the analysis of MAR missing values using the FIML method. In terms of the stability of the reliability estimation, all simulated conditions showed convergence rates under 50%. The lowest convergence rate was 72.9% under the condition of 750 samples with 30% missing values and the large (.296) ICC values fixing the same between-group parameters (Condition #81) when estimating ω_w and ω_B. Among the four cases of the same conditions (i.e., 750 samples with 30% missing values and large (.296) ICC with the same between-group level parameters), the convergence rate under MAR using the LD method showed the lowest (39.8%). Thus, the results showed that (1) MAR missing values caused more severe nonconvergence problems in the two-level MCFA model, and (2) the LD method performed worse than the FIML method in dealing with the missing values.

As the missing data proportions increased from 0% to 30%, the percentage biases of all six reliability measures (single-level coefficient alpha α, level-specific coefficient alpha including within-group level coefficient alpha α_w and between-group level coefficient alpha α_B, single-level coefficient omega ω, level-specific coefficient omega including within-group level coefficient omega ω_w and between-group level coefficient omega ω_B) increased. For example, under the condition of 750 samples and the small (.050) ICC fixing the same between-group level parameters (i.e., Conditions #73–#75), the absolute values of the percentage bias increased from 1.5% to 2.3% for α_B and from 3.5% to 3.7% for ω_B. Thus, the results indicated that the larger the missing data proportions, the larger the inaccuracy in the reliability estimates.

Comparing the effects of the sample sizes on the reliability estimates, when the sample sizes increased from 750 to 1,500 with 100 clusters and 15 cluster sizes (i.e., Conditions #75 versus #84), the percentage biases of the between reliabilities decreased from 2.3% to 0.2% for α_B, and from 3.7% to 1.4% for ω_B. And the percentage biases of α and α_w were 1.5%, while the percentage biases of ω and ω_w were 0%. Also, comparing Condition #75 with Condition #93 (i.e., another 1,500 samples with 50 clusters and 30 cluster sizes), the percentage bias of α_B decreased from 2.3% to 0.6%, and the percentage bias of ω_B decreased from 3.7% to 1.0%.

Thus, the increase of the sample sizes had a positive effect on the accuracy of the reliability estimation, especially for the between-group level reliabilities, α_B and ω_B.

The percentage biases of the reliability estimates changed with the values of ICCs as well. Under the conditions of the same between-group level parameters of 750 samples with 30% missing values when increasing the ICCs from .050 to .296 (i.e., Conditions #75, #78, #81), the percentage biases of the single-level and the within-group level reliability estimates increased (α from 1.5% to 2.2%, α_w from 1.5% to 2.8%, ω from 0.1% to 0.6%, ω_w from 0% to 1.3%), while the percentage bias of the between-group level reliability estimates deceased (α_B from 2.3% to 1.1%, ω_B from 3.7% to 0.3%). Also, under the conditions of the same within-group level parameters of 750 samples with 30% missing values as ICC increased from .050 to .296 (i.e., Conditions #111, #114, #117), the percentage biases of α (1.5% to 2.2%), and α_w (1.5% to 2.0%) increased, while the percentage biases of α_B (2.3% to 0.8%) and ω_B (2.3% to 2.0%) decreased. Thus, the increase of the ICC values reduced the accuracy of the single-level and the within-group level reliability estimates, while the increase of the ICC increased the accuracy of the between-group level reliability indices.

Most of the simulated conditions showed percentage biases of the reliability measures smaller than 15%. Comparing the results of MAR using the LD method, we found that the FIML method outperformed the LD method in terms of the estimation accuracy because of the smaller percentage biases of all reliability indices. However, the percentage biases of all six reliability estimates under MAR were larger than the percentage biases under MCAR. Thus, MAR had a more negative impact on the reliability estimations than MCAR did. By comparing the two types of the reliability estimates, we note that the percentage biases of the single-level and the within-group level coefficient alpha α and α_w were larger than the percentage biases of the single-level and the within-group level coefficient omega ω and ω_w under the simulated conditions of MAR when using the FIML method. However, the percentage bias of α_B was smaller than of ω_B under most of the simulated conditions of MAR when implementing the FIML method.

Results of MNAR Missing Values: LD Method

Figures 18–20 and Tables 24–27 show the convergence rates and the percentage biases of the coefficient alpha, and the coefficient omega estimated from the analysis of MAR missing values using the LD method. In terms of the stability of the reliability estimation, most of the conditions showed over 50% convergence rates. However, under the condition of 750 samples with 30% missing values and the large (.296) ICC with fixing the same between-group level parameters (Condition #153), the two-level MCFA model to estimate the within- and the between-group omega reported less than 50% convergence rates (35.5%). Comparing the convergence rates of the same condition using the LD method among MCAR (Condition #9), MAR (Condition #81) and MNAR (Condition #153) cases, we note that Condition #153 showed the lowest value. Thus, the missing values caused by the MNAR mechanism caused a more serious problem in reliability estimation in terms of the model convergency.

As the missing data proportions increased from 0% to 30%, the percentage biases of all six reliability measures (single-level coefficient alpha α, level-specific coefficient alpha including within-group level coefficient alpha α_w and between-group level coefficient alpha α_B, single-level coefficient omega ω, level-specific coefficient omega including within-group level coefficient omega ω_w and between-group level coefficient omega ω_B) increased. For example, under the condition of 750 samples with the small (.050) ICC fixing the same between-group level parameters when increasing the missing data proportions from 0% to 30% (Conditions #145 to #147), the percentage biases increased from 1.5% to 11% for α and α_w, and from 0% to

11% for ω and ω_w. By the increase of the missing data proportions, the between-group level reliabilities increased to more than 15%: from 1.5% to 29.2% for α_B, and from 3.5% to 25.1% for ω_B. Thus, the increase of the missing data proportions had a negative impact on the accuracy of the reliability estimation, and the impact was more serious for the estimation accuracy of the between-group level reliabilities.

Among the conditions of the sample sizes 750, 1,500, and 3,000, the percentage biases of all reliability estimates increased. For example, comparing the results of 750 samples and 1,500 samples (100 clusters and 15 cluster sizes) under the same conditions of 30% missing values and the small (.050) ICC fixing the same between-group level parameters (i.e., Condition #147 versus Condition #156), the percentage bias of α_B increased from 29.2% to 35.0%, while the percentage biases of α (11.5%) and α_w (11.0%) were not different. And comparing Conditions #147 and #156, the percentage bias increased from 11.9% to 12.0% for ω and from 25.1% to 33.8% for ω_B, and the percentage bias of ω_w (11.3%) was not different between the two conditions. Also, comparing Condition #147 with Condition #165 (i.e., another 1,500 samples with 50 clusters and 30 cluster sizes), the percentage biases increased from 11.5% to 11.6% for α, from 29.2% to 36.0% for α_B, from 11.9% to 12.0% for ω, and from 25.1% to 36.6% for ω_B. And the percentage biases of α_w (11.0%) and ω_w (11.3%) were not different between the two conditions.

Under the conditions fixing the between-group level parameters as the same, when increasing ICCs, the percentage biases of the single-level and the within-group level reliabilities increased, while the percentage biases of the between-group level reliabilities decreased. For example, comparing the conditions of 750 samples with 30% missing values (i.e., Conditions #147, #150, #153), the percentage biases of α increased from 11.5% to 24.5%, α_w increased from 11.1% to 21.8%, ω increased from 11.9% to 25.0%, and ω_w increased from 11.3% to 20.6%. However, under the same conditions, the percentage biases decreased from 29.2% to 14.3% for α_B, and from 25.1% to 7.8% for ω_B.

Contrary to the results of the conditions of the same between-group level parameters, under the conditions of the same within-group level parameters, the percentage biases of the single-level and the between-group level reliability indices increased, while the percentage bias of the within-group level reliability indices decreased. For example, among three conditions of 750 samples with 30% missing values (i.e., Conditions #183, #186, #189), the percentage biases increased from 11.5% to 14.2% for α, from 29.2% to 48.9% for α_B, from 11.9% to 14.8% for ω, and from 25.1% to 38.6% for ω_B. However, the percentage biases of α_w decreased from 11.0% to 8.3% and of ω_w from 11.3% to 10.3%. Thus, the increase of the ICCs reduced the accuracy of the single-level reliability estimates α and ω, whereas the increase of the ICCs had a positive impact on the between-group level reliability estimates α_B and ω_B.

Regarding the cutoff criterion for percentage bias, the single-level and the within-group level reliability estimates showed more than 15% bias when the large (.296) ICC fixing the same between-group level parameters with above 15% missing data proportions. And under the conditions of 30% missing data proportions with the medium (.111) fixing the same within-group level parameters, the percentage biases of all six reliability estimates were larger than 15%. And the between-group level reliability estimates showed over 15% bias, when the small (.050) ICC with above 15% missing data proportions.

Comparing the two types of reliability estimates, the percentage bias of the coefficient alpha was larger than the percentage bias of the coefficient omega under both the single-level CFA model and the two-level MCFA model. And among three missing data mechanisms, the percentage bias under MNAR showed the largest values when using the LD method. The results indicated that the LD method produced more biased results when the missing data were generated by the MNAR mechanism.

Results of MNAR Missing Values: FIML Method

Figures 21–23 and Tables 28–31 show the convergence rates and the percentage biases of the coefficient alpha, and the coefficient omega estimated from the analysis of MNAR missing values utilizing the FIML method. In terms of the stability of the reliability estimation, all simulated conditions showed convergence rates under 50%. The lowest convergence rate was 57.4% under the condition of 750 samples with 30% missing values and the large (.296) ICC fixing the same within-group level parameters (Condition #153). Because the conditions of the small samples (750 with 50 clusters and 15 cluster sizes) with the large (30%) missing data proportions and the large (.296) ICC fixing the same within-group level parameters showed the lowest convergence rates in each of the cases manipulating the missing data mechanism (either MCAR, MAR, or MNAR) and the missing data techniques (either the LD or the FIML method), the combinations of these factors had a negative impact on the convergence of the models in estimating the single-level and the within- and the between-group level reliabilities. Also, the convergence rate under the MAR using the LD method (39.8%) and the MNAR using LD method (35.5%) showed less than 50% convergence rates. Thus, the use of the LD method to deal with either MAR or MNAR missing data created the convergency issue in the two-level MCFA model in estimating the within- and the between-group level omega, ω_W and ω_B.

In terms of the accuracy of the reliability estimation, we first compared the percentage biases regarding the missing data proportions. When we increased the missing data proportions from 0% to 30%, the absolute values of the percentage biases of all six reliability estimates increased under all simulated conditions. For example, among Conditions #145–#147 (i.e., the conditions of 750 samples with the small (.050) ICC fixing the same between-group level parameters), the percentage biases of α and α_w increased from 1.5% to 4.6%, and the percentage bias of α_B increased from 1.5% to 8.2%. And under the same conditions, the percentage biases of ω and ω_w increased from 0% to 3.5%, and the percentage bias of ω_B increased from 3.5% to 7.7%. The results indicated that the increase of the missing data proportions reduced the accuracy of the reliability estimation, especially for the between-group level reliabilities.

When the sample sizes increased, the percentage biases of all six reliability estimates decreased in most of the simulated conditions under MNAR using the FIML method. For example, comparing the conditions of 750 samples to 1,500 samples with 100 clusters and 15 cluster sizes fixing the small ICC (.050) and the same between-group level parameters (i.e., Conditions #147 versus #156), the percentage bias decreased from 8.2% to 5.7% for α_B and from 7.7% to 5.7% for ω_B. And the percentage biases of α (4.6%), α_w (4.6%), ω (3.6%), and ω_w (3.5%) did not change between the two conditions. In addition, comparing Condition #147 with Condition #165 (i.e., another condition of 1,500 samples with 50 clusters and 30 cluster sizes), the percentage biases decreased from 4.6% to 4.5% for α_w, from 8.2% to 5.4% for α_B, and from 7.7% to 4.8% for ω_B.

As ICCs increased, the percentage biases of the single-level and the within-group level reliability indices increased, while the percentage biases of the between-group level reliabilities decreased. Under the conditions of 750 samples with 30% missing values fixing the same between-group level parameters, when the ICCs increased from .050 to .296 (i.e., Conditions #147, #150, #153), the percentage biases increased from 4.6% to 17.4% for α, from 4.6% to 20.8% for α_w, from 3.6% to 18.0% for ω, and from 3.5% to 21.3% for ω_w. However, the percentage biases of α_B (8.2% to 2.2%) and ω_B (7.7% to 0.2%) decreased by increasing the ICCs. In addition, under the conditions of 750 samples with 30% missing values fixing the same within-group level parameters, as ICCs increased the percentage biases of α (4.6% to 6.0%) and ω (3.6% to 6.5%) increased, while the percentage biases of α_B (8.2% to 6.6%) and ω_B (7.6% to 5.1%) decreased. Thus, the increase of ICCs had a positive impact on the accurate estimation of the between-group level reliabilities. However, the increase of ICCs reduced the accuracy of the estimation of the single-level and the within-group level reliability indices.

Regarding the cutoff criterion for the percentage bias, most of the simulated conditions showed the percentage biases of the reliability measures smaller than 15%. However, under the conditions of the same between-group level parameters, when the ICC was large (.296) and the missing data proportions were 30%, the percentage biases of the single-level and the within-group level coefficient alpha and omega were larger than 15%, while the between-group level coefficient alpha and omega were smaller than 15%. However, under the conditions of the same within-group level parameters, all six reliability measures showed percentage biases less than 15% regardless of the magnitudes of ICCs, sample sizes, and missing data proportions.

A comparison of the two types of the reliability indices showed that the percentage bias of the coefficient alpha was larger than the percentage bias of the coefficient omega when dealing with MNAR in using the FIML method. Also, the percentage biases of all six reliability estimates when using the FIML method were smaller than when utilizing the LD method. In addition, the results of MNAR were larger than other two missing data mechanisms, MCAR and MAR. Thus, the existence of MNAR missing values and the usage of the LD method had the most negative impact on the estimation of the reliability indices, especially for the coefficient alpha.

CHAPTER 5

CONCLUSIONS AND DISCUSSIONS

In this study we investigated the reliability estimates regarding the conditions of the missing values and the multilevel data structure. The empirical data analysis showed that because the single-level coefficient alpha α, and the within-group and between-group level coefficient alpha α_W and α_B were estimated from the fully saturated model, the results showed a perfect model fit. However, the single-level coefficient omega ω and the within-group and between-group level coefficient omega ω_W and ω_B showed convergence issues that did not satisfy good model fit results (i.e., CFI > .90, TLI > .90, and RMSEA < .08).

To compare the reliability estimates, except for ω_B of ST038, the between-group level reliability estimates were around .90, while the single-level and the within-group level reliability estimates were around .80. For ST038, the largest difference between each reliability index of the two missing data techniques (i.e., listwise deletion [LD] method and full information maximum likelihood [FIML] method) was the change of ω_B estimates: .079. However, for IC013, the difference between the reliability estimates of the LD and FIML methods was minimal, from .002 to .004. This was because the intra-class correlation (ICC) value of ST038 was smaller (.003 using the LD method, .005 using the FIML method) than the ICC value of IC013 (.029 using the LD and FIML methods). Geldhof et al. (2014) showed that when the ICC is small, the between-group level reliability estimates show more bias. Even though we did not know the population values of the reliability estimates, the large changes of ω_B indicated the possibility of a large bias in the between-group level reliability estimation. In addition, the difference between the reliability estimates of the LD and FIML methods corresponded to the findings of the previous studies (Cuesta Izquierdo & Fonseca Pedrero, 2014; Downey & King, 1998; Enders, 2003, 2004; McDonald et al., 2000; Raykov, 2009b), indicating that the usage of the missing data techniques had a significant impact on the reliability estimation.

In the simulation study, we manipulated the conditions not only for the ICCs (.050 as small, .111 as medium, and .296 as large) and missing data techniques (LD and FIML methods), but also for the missing data mechanisms (missing completely at random [MCAR], missing at random [MAR], and missing not at random [MNAR]), missing data proportions (0%, 15%, and 30%), and sample sizes (750 samples, 1,500 samples with 50 clusters, 1,500 samples with 100 clusters, and 3,000 samples).

In terms of convergence rates, because the single-level coefficient alpha and the within-group and between-group level coefficient alpha were estimated from the fully saturated model, the results showed a perfect model fit (i.e., 100% convergence rates) under all simulated conditions. For the coefficient omega estimates, the convergence rates were below 50% (39.8% under MAR, 35.5% under MNAR) when (a) we used the LD method, (b) the sample size was small (750 samples), and (c) the ICC value was large (.296) by fixing the same between-group level parameters, the two-level MCFA model to estimate the within-group and the between-group coefficient omega ω_W and ω_B.

Hancock and An (2020) showed that the coefficient omega estimates tended toward nonconvergence under the conditions of smaller sample sizes and the low population values of the reliability index. From Raykov (2004), the low reliability indicates that the results of the test would be inconsistent depending on the features of the samples or the test administration conditions. In our simulation study, the condition showing the lowest convergence rates was the combination of the negative factors for the convergence rates of the coefficient omega calculated by Hancock and An (2020): (a) the sample size was 750, which was the smallest value in the sample sizes conditions, (b) after deleting the missing values using the LD method, the sample sizes were below 750 samples, (c) the large ICC conditions with the same between-group parameters set the smallest population values of the within-group level coefficient omega (ω_w=.376). Thus, the results of the lowest convergence rates supported the previous findings of Hancock and An (2020).

In terms of the percentage bias, the quantitative methodological findings from the simulation study were as follows. The single-level and within-group level reliability estimates reported the highest percentage bias under the conditions of (a) the small sample size (750 samples), (b) the large ICC value (.296), and (c) the low population values of the within-group reliabilities (α_w=.372, ω_w=.376). Meanwhile the between-group level reliability estimates provided the worst accuracy results under the conditions of (a) the small samples (750), (b) the small ICC value (.050), and (c) the low population values of the between-group level reliabilities (α_B=.372, ω_B=.376). The significant effects of the ICCs and the population values of the reliabilities corresponded to the findings of Geldhof et al. (2014). That is, when there is a discrepancy of the reliabilities at the within-group level and the between-group level, then, the estimates result of the multilevel-reliabilities would be biased.

And under most simulated conditions in this study, the percentage bias of the coefficient alpha was higher than the percentage bias of the coefficient omega. In the data generation step of the simulation study, we set the population values of the factor loadings and the error variances varied, because it is a case of the congeneric measurement model (i.e., the assumptions of the different factor loadings and the different error variances are varied among the indicators;

Jöreskog, 1971) which is more general in the practical testing situation (Cortina, 1993; Raykov, 1997a, 1997b; Sijtsma, 2009; Yang & Green, 2011; Zimmermanet al., 1993). However, the conditions to set the different factor loadings and the error variances violate the assumptions of tau-equivalent measurement model that the single-level, the within-group and the between-group level coefficient alpha estimates are based on. Thus, the results of this study supported the findings of the previous studies that the violation of the tau-equivalent assumptions caused the overestimated values of the single-level coefficient alpha α (Deng & Chan, 2017, Duun et al., 2014; Green & Yang, 2009a; Raykov & Marcoulides, 2015), and the level-specific coefficient alpha α_w, α_B estimates (Bonito et al., 2012; Geldhof et al., 2014).

In addition, regarding the conditions of the missing values, if (a) the missing data were generated by MNAR, (b) the missing data were dealt with by the LD method, and (c) the proportion of the missing values was 30%, the percentage bias of the reliability estimates was high. Rubin (1976) showed that MNAR generates the non-ignorable missing values that cause the systematic loss of the entire data set. In addition, the large (30%, about 1/3) proportions of the missing values and the usage of the LD method (i.e., a complete case method; a way to remove the cases that have at least one missing value) make the systematic loss of the data set more severe. That is, the combinations of the large proportions of MNAR missing values with the LD method causes the data set to be less representative of the population of interest. Thus, the results would be biased in the statistical inference of the data analysis, such as multiple regression model (Enders, 2001b), latent variable analysis (Cai & Song, 2010; Enders, 2011; Lu et al., 2011; Muthén et al., 1987), item response theory model (Rose et al., 2010), or the computerized adaptive testing model (Cetin-Berber et al., 2019).

For the reliability estimation, Enders (2003) calculated that under the conditions of 30% MNAR, missing data utilizing the LD method showed the highest root-mean-square error value (RMSE) compared with the results under the conditions of 15% MCAR or MAR missing values using the expectation-maximization (EM) algorithm method or multiple imputation method. The results of the percentage bias of MNAR in this study extended the findings of Enders (2003); the way to deal with MNAR missing data (above 30%) by using the LD method would not be a good choice to estimate both the single-level reliability and the within-group and between-group reliability indices.

For the practical researchers and applied methodologists, this study suggests the specific guidelines to report the reliability of the measurement tool in actual testing situations regarding the data features. Specifically, if above 30% of the data are missing, it would be better to use the FIML method to estimate the reliability indices. In addition, when the sample sizes are around 750, researchers should check the ICC values to determine which reliability indices would be used; if the ICC is high (about .30), the between-group level reliability index is a better choice, and if the ICC is approximately .05, using the single-level or within-group level reliability indices is preferable. Kim et al. (2016), showed that reporting the reliability of the items analyzed by the multilevel models is not common because of the lack of relevant references; therefore, the results of this study would be one of the references used.

The findings of this study suggest that future research could be conducted in the following directions. First, more conditions of the missing data could be examined, such as utilizing the other missing data techniques, especially for the data imputation methods. The previous scholars investigating the performance of the single-level reliability compared the results of the data deletion method with the data imputation methods (Enders, 2003; McDonald et al., 2000) or the performance among several data imputation methods (Downey & King, 1998), such as person mean imputation, item mean imputation, or multiple linear regression imputation. Even though the multilevel data structure was not considered, Enders (2004) showed that the method using the EM algorithm (i.e., one of the data augmentation methods, such as the FIML method) performed better than the data deletion (LD and pairwise deletion) methods and the item mean imputation method to estimate the single-level coefficient alpha. Thus, the inclusion of the data imputation methods into the simulation study design would be a testable condition for comparing the reliability estimates with missing values.

Second, regarding the conditions of the multilevel data structure, we could examine more data conditions, such as the number of items (i.e., the test length), the number of answer categories (i.e., the point-scale of the items), or the complexity of the model. For the multilevel reliability indices, Bonito et al. (2012) showed that under the condition of five items in the test, the reliability of ignoring the data hierarchy showed more inflated values than the individual-level and group-level reliability estimates. Further, Lai (2021) showed that the within-group and between-group level coefficient omega could be inflated under the more complicated models that had multiple factors, or the factors are correlated to each other, rather than the single-factor model we examined. And Zinbarg et al. (2005) examined the misspecification caused the bias in the estimation of the coefficient omega. In addition, Kim et al. (2020) proposed the generalized nonlinear SEM reliability coefficient based on the factor analytic framework, which showed a promising performance to deal with the test items having an uneven (i.e., either two or five) number of answer categories. Hence, in addition to the sample sizes and the ICCs examined in this study, the effects of other data features, such as the number of items (i.e., the test length), the number of the answer categories (i.e., the point-scale of the items), the complexity of the model, or the misspecification of the model could be considered in the future research.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks-Cole.

Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological Methodology, 17,* 71–103. https://doi.org/271029

Allison, P. D. (2003). Missing Data Techniques for Structural Equation Modeling. *Journal of Abnormal Psychology, 112*(4), 545–557. https://doi.org/10.1037/0021-843X.112.4.545

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association. Retrieved from http://www.apa.org/science/programs/testing/standards.aspx

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*(1), 5–37. https://doi.org/10.1016/j.jsp.2009.10.001

Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1*(4), 343–357. https://doi.org/10.1016/0049-089X(72)90082-8

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Black, A. C., Harel, O., & Betsy McCoach, D. (2011). Missing data techniques for multilevel data: Implications of model misspecification. *Journal of Applied Statistics, 38*(9), 1845–1865. https://doi.org/10.1080/02664763.2010.529882

Bollen, K. A. (1989). *Structural equations with latent variables*. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781118619179

Bonito, J. A., Ruppel, E. K., & Keyton, J. (2012). Reliability estimates for multilevel designs in group research. *Small Group Research, 43*(4), 443–467. https://doi.org/10.1177/1046496412437614

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cuttance & R. Ecob (Eds.), *Structural modeling by example* (pp. 160–188). New York, NY: University of Cambridge.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147–158. https://doi.org/10.1037/0033-2909.101.1.147

Cai, J. H., & Song, X. Y. (2010). Bayesian analysis of mixtures in structural equation models with non-ignorable missing data. *British Journal of Mathematical and Statistical Psychology, 63*(3), 491–508. https://doi.org/10.1348/000711009X475187

Cai, J. H., Song, X. Y., & Hser, Y. I. (2010). A Bayesian analysis of mixture structural equation models with non-ignorable missing responses and covariates. *Statistics in Medicine, 29*(18), 1861–1874. https://doi.org/10.1002/sim.3915

Cetin-Berber, D. D., Sari, H. I., & Huggins-Manley, A. C. (2019). Imputation methods to deal with missing responses in computerized adaptive multistage testing. *Educational and Psychological Measurement, 79*(3), 495–511. https://doi.org/10.1177/0013164418805532

Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research, 52*(1), 12–30. https://doi.org/10.1080/00273171.2016.1245600

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research, 84*(4), 487–508. https://doi.org/10.3102/0034654314532697

Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, *19*(4), 651–682. https://doi.org/10.1177/1094428116656239

Cho, S.-J., & Preacher, K. J. (2016). Measurement error correction formula for cluster-level group differences in cluster randomized and observational studies. *Educational and Psychological Measurement*, *76*(5), 771–786. https://doi.org/10.1177/0013164415612255

Cokluk, O., & Kayri, M. (2011). The Effects of Methods of Imputation for Missing Values on the Validity and Reliability of Scales. *Educational Sciences: Theory and Practice, 11*(1), 303–309. https://www.proquest.com/scholarly-journals/effects-methods-imputation-missing-values-on/docview/861327961/se-2?accountid=14537

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart and Winston.

Cronbach, L. J. (1943). On estimates of test reliability. *Journal of Educational Psychology, 34*(8), 485–494. https://doi.org/10.1037/h0058608

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391–418. https://doi.org/10.1177/0013164404266386

Cuesta Izquierdo, M., & Fonseca Pedrero, E. (2014). Estimating the reliability coefficient of tests in presence of missing values. *Psicothema, 26*(4), 516–523. https://doi.org/10.7334/psicothema2014.98

Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, *77*(2), 185–203. https://doi.org/10.1177/0013164416658325

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publications, Inc.

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus, 2*(1), 1–17. https://doi.org/10.1186/2193-1801-2-222

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *Journal of General Psychology*, *125*(2), 175–191. https://doi.org/10.1080/00221309809595542

Duhachek, A., & Iacobucci, D. (2004). Alpha’s standard error (ASE): an accurate and precise confidence interval estimate. *Journal of Applied Psychology*, *89*(5), 792–808. https://doi.org/10.1037/0021-9010.89.5.792

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Edwards, A. A., Joyner, K. J., & Schatschneider, C. (2021). A simulation study on the performance of different reliability estimation methods. *Educational and Psychological Measurement*, *81*(6), 1089–1117. https://doi.org/10.1177/0013164421994184

Enders, C. K. (2001a). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352–370. https://doi.org/10.1037/1082-989X.6.4.352

Enders, C. K. (2001b). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, *61*(5), 713–740. https://doi.org/10.1177/00131640121971482

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, *8*(3), 322–337. https://doi.org/10.1037/1082-989X.8.3.322

Enders, C. K. (2004). The impact of missing data on sample reliability estimates Implications for reliability reporting practices. *Educational and Psychological Measurement*, *64*(3), 419–436. https://doi.org/10.1177/0013164403261050

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, *16*(1), 1–16. https://doi.org/10.1037/a0022640

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation*

Modeling: A Multidisciplinary Journal, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Enders, C., Dietz, S., Montague, M., & Dixon, J. (2006). Modern Alternatives for Dealing with Missing Data in Special Education Research. *Advances in Learning and Behavioral Disabilities, 19*, 101–129. https://doi.org/10.1016/S0735-004X(06)19005-9

Estabrook, R., & Neale, M. (2013). A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to nicotine dependence. *Multivariate Behavioral Research*, *48*(1), 1–27. https://doi.org/10.1080/00273171.2012.730072

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). Washington, DC: American Council on Education.

Feldt, L. S., & Qualls, A. L. (1996). Bias in coefficient alpha arising from heterogeneity of test content. *Applied Measurement in Education*, *9*, 277–286. https://doi.org/10.1207/s15324818ame0903_5

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*(4), 409–420. https://doi.org/10.1007/BF02296204

Garciarena, U., & Santana, R. (2017). An extensive analysis of the interaction between missing data types, imputation methods, and supervised classifiers. *Expert Systems with Applications, 89*, 52–65. https://doi.org/10.1016/j.eswa.2017.07.026

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91. https://doi.org/10.1037/a0032138

Goldstein, H., Carpenter, J. R., & Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *177*(2), 553–564. https://doi.org/10.1111/rssa.12022

Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*(1), 121–135. https://doi.org/10.1007/s11336-008-9098-4

Green, S. B., & Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*(4), 827–838. https://doi.org/10.1177/001316447703700403

Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods, 21*(1), 111–149. https://doi.org/10.1177/1094428117703686

Grund, S., Lüdtke, O., & Robitzsch, A. (2019). Missing data in multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 365–386). Washington: American Psychological Association.

Hancock, G. R., & An, J. (2020). A closed-form alternative for estimating ω reliability under unidimensionality. *Measurement: Interdisciplinary Research and Perspectives*, *18*(1), 1–14. https://doi.org/10.1080/15366367.2019.1656049

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, *60*, 523–531. http://dx.doi.org/10.1177/00131640021970691

Hox, J. J. (2010). *Multilevel analysis techniques and applications* (2nd ed.). New York, NY: Routledge.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, *6*(3), 153–160. https://doi.org/10.1007/BF02289270

Hsu, H. Y., Lin, J. H., Kwok, O. M., Acosta, S., & Willson, V. (2017). The impact of intraclass correlation on the effectiveness of level-specific fit indices in multilevel structural equation modeling: A Monte Carlo study. *Educational and Psychological Measurement, 77*(1), 5–31. https://doi.org/10.1177/0013164416642823

Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test*, *18*(1), 1–43. https://doi.org/10.1007/s11749-009-0138-x

Jackson, P., & Agunwamba, C. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items: I: Algebraic lower bounds. *Psychometrika, 42*(4), 567–578. https://doi.org/10.1007/BF02295979

Jak, S., & Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Frontiers in Psychology*, *8*, 1640. https://doi.org/10.3389/fpsyg.2017.01640

Jeon, M. J., Lee, G., Hwang, J. W., & Kang, S. J. (2009). Estimating reliability of school-level scores using multilevel and generalizability theory models. *Asia Pacific Education Review*, *10*(2), 149–158. https://doi.org/10.1007/s12564-009-9014-3

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*(2), 109–133. https://doi.org/10.1007/BF02291393

Kaitz, H. (1945). A note on reliability. *Psychometrika, 10*(2), 127–131. https://doi.org/10.1007/BF02288879

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, *23*(1), 69–86. https://doi.org/10.1207/s15327906mbr2301_4

Kelley, T. (1942). The reliability coefficient. *Psychometrika, 7*(2), 75–83. https://doi.org/10.1007/BF02288068

Kim, E. S., & Cao, C. (2015). Testing group mean differences of latent variables in multilevel data using multiple-group multilevel CFA and multilevel MIMIC modeling. *Multivariate Behavioral Research*, *50*(4), 436–456. https://doi.org/10.1080/00273171.2015.1021447

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research, 51*(6), 881–898. https://dx.doi.org/10.1080/00273171.2016.1228042

Kim, S., Lu, Z., & Cohen, A. S. (2020). Reliability for tests with items having different numbers of ordered categories. *Applied Psychological Measurement, 44*(2), 137–149. https://doi.org/10.1177/0146621619835498

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.

Koch, T., Schultze, M., Burrus, J., Roberts, R. D., & Eid, M. (2015). A multilevel CFA-MTMM model for nested structurally different methods. *Journal of Educational and Behavioral Statistics, 40*(5), 477–510. https://doi.org/10.3102/1076998615606109

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2,* 151–160. https://doi.org/10.1007/BF02288391

Lai, M. H. C. (2021). Composite reliability of multilevel data: It's about observed scores and construct meanings. *Psychological Methods*, *26*(1), 90–102. https://doi.org/10.1037/met0000287

Leite, W., & Beretvas, S. N. (2010). The performance of multiple imputation for Likert-type items with missing data. *Journal of Modern Applied Statistical Methods, 9*(1), 8. https://doi.org/10.22237/jmasm/1272686820

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods, 1*(1), 98–107. https://doi.org/10.1037/1082-989X.1.1.98

Li, J., & Lomax, R. G. (2017). Effects of missing data methods in SEM under conditions of incomplete and nonnormal data. *The Journal of Experimental Education*, *85*(2), 231–258. https://doi.org/10.1080/00220973.2015.1134418

Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, *46*(4), 567–597. https://doi.org/10.1080/00273171.2011.589261

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, *22*(1), 141–165. https://doi.org/10.1037/met0000096

Marcoulides, G. A., & Schumacker, R. E. (2013). *Advanced structural equation modeling: Issues and techniques* (1st ed.). New York, NY: Psychology Press.

Mcdonald, R. A., Thurston, P. W., & Nelson, M. R. (2000). A Monte Carlo study of missing item methods. *Organizational Research Methods, 3*(1), 71–92. https://doi.org/10.1177/109442810031003

McDonald, R. P. (1978). Generalizability in Factorable Domains:" Domain Validity and Generalizability". *Educational and Psychological Measurement*, *38*(1), 75–79. https://doi.org/10.1177/001316447803800111

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction*. New York, NY: Guilford Press.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education.

Miller, M.B (1995). Coefficient Alpha: A basic intro from the perspective of classical test theory and SEM. *Structural Equation Modeling: A Multidisciplinary Journal, 2*(3). 255–273. https://doi.org/10.1080/10705519509540013

Mulaik, S. (2009). *Foundations of Factor Analysis* (2nd ed.). New York, NY: Chapman and Hall/CRC.

Muthén, B. O. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods & Research, 22*(3), 376–398. https://doi.org/10.1177/0049124194022003006

Muthén, B. O. (1997). 10. Latent variable modeling of longitudinal and multilevel data. *Sociological Methodology, 27*(1), 453–480. https://doi.org/10.1111/1467-9531.271034

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*(3), 431–462. https://doi.org/10.1007/BF02294365

Muthén, L., & Muthén, B. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Novick, M. R., & Lewis, C. (1966). Coefficient alpha and the reliability of composite measurements. *ETS Research Bulletin Series, 1966*(1), i–28. https://doi.org/10.1002/j.2333-8504.1966.tb00356.x

Padgett, R. N., & Morgan, G. B. (2021). Multilevel CFA with ordered categorical data: A simulation study comparing fit indices across robust estimation methods. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 51-68. https://doi.org/10.1080/10705511.2020.1759426

Padilla, M. A., & Divers, J. (2013). Bootstrap interval estimation of reliability via coefficient omega. *Journal of Modern Applied Statistical Methods, 12*(1), 13. https://doi.org/10.22237/jmasm/1367381520

Padilla, M. A., & Divers, J. (2016). A comparison of composite reliability estimators: coefficient omega confidence intervals in the current literature. *Educational and Psychological Measurement, 76*(3), 436–453. https://doi.org/10.1177/0013164415593776

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*(4), 525–556. https://doi.org/10.3102/00346543074004525

Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: Some consequences for model fit and standardized parameter estimates. *Multivariate Behavioral Research, 49*(6), 518–543. https://doi.org/10.1080/00273171.2014.933762

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). London, England: Sage.

Raykov, T. (1997a). Scale reliability, Cronbach's Coefficient Alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329–353. https://doi.org/10.1207/s15327906mbr3204_2

Raykov, T. (1997b). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35*(2), 299–331. https://doi.org/10.1016/S0005-7894(04)80041-8

Raykov, T. (2009a). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development, 42*(3), 223–232. https://doi.org/10.1177/0748175609344096

Raykov, T. (2009b). Interval estimation of revision effect on scale reliability via covariance structure modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 539–555. https://doi.org/10.1080/10705510903008337

Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472–492). New York, NY: Guilford.

Raykov, T., & Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *British Journal of Mathematical and Statistical Psychology, 58*(1), 65–82. https://doi.org/10.1348/000711005X38753

Raykov, T., & Marcoulides, G. A. (2006). On multilevel model reliability estimation from the perspective of structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(1), 130–141. https://doi.org/10.1207/s15328007sem1301_7

Raykov, T., & Marcoulides, G. A. (2015). A direct latent variable modeling based method for point and interval estimation of coefficient alpha. *Educational and Psychological Measurement, 75*(1), 146–156. https://doi.org/10.1177/0013164414526039

Revelle, W. (1979). Hierarchical cluster-analysis and the internal structure of tests. *Multivariate Behavioral Research, 14*(1), 57–74. https://doi.org/10.1207/s15327906mbr1401_4

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145–154. https://doi.org/10.1007/s11336-008-9102-z

Rockel, T. (2020). *missMethods*: Methods for Missing Data. R package version 0.2.0. https://CRAN.R-project.org/package=missMethods

Rose, N., Von Davier, M., & Xu, X. (2010). Modeling nonignorable missing data with item response theory (IRT). *ETS Research Report Series, 2010*(1), i–53. https://doi.org/10.1002/j.2333-8504.2010.tb02218.x

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537–560. https://doi.org/10.1111/j.1744-6570.1994.tb01736.x

Rousseau, M., Simon, M., Bertrand, R., & Hachey, K. (2012). Reporting missing data: a study of selected articles published from 2003–2007. *Quality & Quantity, 46*(5), 1393–1406. https://doi.org/10.1007/s11135-011-9452-y

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 583–601. https://doi.org/10.1080/10705510903203466

Santos, M. S., Pereira, R. C., Costa, A. F., Soares, J. P., Santos, J., & Abreu, P. H. (2019). Generating synthetic missing data: A review by missing mechanism. *IEEE Access, 7*, 11651–11667. https://doi.org/10.1109/ACCESS.2019.2891360

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Sideridis, G., Saddaawi, A., & Al-Harbi, K. (2018). Internal consistency reliability in measurement Aggregate and multilevel approaches. *Journal of Modern Applied Statistical Methods, 17*(1), 15. https://doi.org/10.22237/jmasm/1530027194

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*(4), 317–329. https://doi.org/10.1037/1082-989X.6.4.317

Snijders, T. A. B., & Bosker, R. J. (1999). Multilevel analysis. Thousand Oaks, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, England: Sage.

Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–293. https://doi.org/10.2307/1412107

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458

Twala, B. (2009). An empirical comparison of techniques for handling incomplete data using decision trees. *Applied Artificial Intelligence, 23*(5), 373–405. https://doi.org/10.1080/08839510902872223

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook for advanced multilevel analysis* (pp. 173–196). Routledge/Taylor & Francis Group.

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). New York, NY: Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Woodhouse, B., & Jackson, P. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika, 42*(4), 579–591. https://doi.org/10.1007/BF02295980

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century?. *Journal of Psychoeducational Assessment, 29*(4), 377–392. https://doi.org/10.1177/0734282911406668

Zhang, Z., & Yuan, K.-H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement, 76*(3), 387–411. https://doi.org/10.1177/0013164415594658

Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement, 53*(1), 33–49. https://doi.org/10.1177/0013164493053001003

Zinbarg, R.E., Revelle, W., Yovel, I., Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ω_H their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7

Tables

Table 1

Test Items of PISA 2018 for Empirical Data Analysis

ST038	
During the past 12 months, how often have you had the following experiences in school?	
Item	Question
ST038_1	Other students left me out of things on purpose.
ST038_2	Other students made fun of me.
ST038_3	I was threatened by other students.
ST038_4	Other students took away or destroyed things that belonged to me.
ST038_5	I got hit or pushed around by other students.
ST038_6	Other students spread nasty rumors about me.
IC013	
Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?	
Item	Question
IC013_1	I forget about time when I'm using digital devices.
IC013_2	The Internet is a great resource for obtaining information I am interested in (e.g., news, sports, dictionary).
IC013_3	It is very useful to have social networks on the Internet.
IC013_4	I am really excited discovering new digital devices or applications.
IC013_5	I really feel bad if no Internet connection is possible.
IC013_6	I like using digital devices.

Table 2

Descriptive Statistics of Six Items of ST038 and IC013 for Empirical Data Analysis

ST038	N_Response				N_Student (%)	N_Missing (%)	Mean (S.D.)
	1* (%)	2* (%)	3* (%)	4* (%)			
ST038_1	2,844 (58.8)	1,151 (23.8)	407 (8.4)	176 (3.6)	4,578 (94.6)	260 (5.4)	1.5 (0.8)
ST038_2	2,651 (54.8)	1,125 (23.3)	465 (9.6)	315 (6.5)	4,556 (94.2)	282 (5.8)	1.7 (0.9)
ST038_3	3,710 (76.7)	552 (11.4)	204 (4.2)	98 (2.0)	4,564 (94.3)	274 (5.7)	1.3 (0.6)
ST038_4	3,871 (80.0)	483 (10.0)	162 (3.3)	57 (1.2)	4,573 (94.5)	265 (5.5)	1.2 (0.6)
ST038_5	3,858 (79.7)	454 (9.4)	174 (3.6)	86 (1.8)	4,572 (94.5)	266 (5.5)	1.2 (0.6)
ST038_6	3,315 (68.5)	772 (16.0)	313 (6.5)	171 (3.5)	4,571 (94.5)	267 (5.5)	1.4 (0.8)
IC013	N_Response				N_Student (%)	N_Missing (%)	Mean (S.D.)
	1** (%)	2** (%)	3** (%)	4** (%)			
IC013_1	399 (8.2)	1,443 (29.8)	1,992 (41.2)	579 (12.0)	4,413 (91.2)	425 (8.8)	2.6 (0.8)
IC013_2	132 (2.7)	390 (8.1)	2,373 (49.0)	1,505 (31.1)	4,400 (90.9)	438 (9.1)	3.2 (0.7)
IC013_3	167 (3.5)	465 (9.6)	2,591 (53.6)	1,150 (23.8)	4,373 (90.4)	465 (9.6)	3.1 (0.7)
IC013_4	202 (4.2)	962 (19.9)	2,321 (48.0)	897 (18.5)	4,382 (90.6)	456 (9.4)	2.9 (0.8)
IC013_5	371 (7.7)	1,573 (32.5)	1,758 (36.3)	701 (14.5)	4,403 (91.0)	435 (9.0)	2.6 (0.8)
IC013_6	126 (2.6)	295 (6.1)	2,432 (50.3)	1,548 (32.0)	4,401 (91.0)	437 (9.0)	3.2 (0.7)

Note. ST038=During the past 12 months, how often have you had the following experiences in school?; IC013=Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?; N_Response=the number of responses: for ST038, 1*=never or almost never, 2*=a few times a year, 3*=a few times a month, 4*=once a week or more, for IC013, 1**=strongly disagree, 2**=disagree, 3**=agree, 4**=strongly agree; N_Student=the number of students excluding the missing values; N_Missing=the number of missing values; Mean=the average of the responses with standard deviations (S.D.).

Table 3

ICC Values, Model-fit Indices, and Reliability Estimates of ST038 and IC013

	LD Method							
	ST038 (ICC=.003)				IC013 (ICC=.029)			
	Est.	CFI	TLI	RMSEA	Est.	CFI	TLI	RMSEA
α	.859	1.000	1.000	.000	.825	1.000	1.000	.000
α_w	.856	1.000	1.000	.000	.821	1.000	1.000	.000
α_B	.940	-	-	-	.928	-	-	-
ω	.847	.906	.843	.159	.808	.962	.936	.091
ω_w	.849	.895	.824	.097	.805	.958	.930	.063
ω_B	.754	-	-	-	.947	-	-	-
	FIML Method							
	ST038 (ICC=.005)				IC013 (ICC=.029)			
	Est.	CFI	TLI	RMSEA	Est.	CFI	TLI	RMSEA
α	.860	1.000	1.000	.000	.823	1.000	1.000	.000
α_w	.858	1.000	1.000	.000	.819	1.000	1.000	.000
α_B	.944	-	-	-	.924	-	-	-
ω	.848	.908	.847	.157	.806	.962	.937	.089
ω_w	.850	.895	.824	.098	.803	.958	.930	.061
ω_B	.833	-	-	-	.944	-	-	-

Note. ST038=During the past 12 months, how often have you had the following experiences in school?; IC013=Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?; Est.= estimates of reliability indices; CFI=comparative fit index, TLI=Tucker-Lewis index, RMSEA=root mean square error of approximation; α=single-level alpha, α_w=within-group level alpha, α_B=between-group level alpha, ω=single-level omega, ω_w=within-group level omega, ω_B=between-group level omega.

Table 4

Population Values for ICCs and Reliabilities by Factor Loadings and Error Variances

(1) Same Between-group Level Parameters, but Different Within-group Level Parameters		
ICC	Reliability	Parameters
.050	$\alpha = .852$	$\alpha = \alpha_w(1 - ICC) + \alpha_B * ICC$
	$\omega = .854$	$\omega = \omega_w(1 - ICC) + \omega_B * ICC$
	$\alpha_w = .852$	$\lambda_{w1} = \lambda_{w2} = .599, \lambda_{w3} = \lambda_{w4} = .699, \lambda_{w5} = \lambda_{w6} = .799$
	$\alpha_B = .850$	$\theta_{w11} = \theta_{w22} = .451, \theta_{w33} = \theta_{w44} = .501, \theta_{w55} = \theta_{w66} = .551$
	$\omega_w = .854$	$\lambda_{B1} = \lambda_{B2} = .137,\ \lambda_{B3} = \lambda_{B4} = .160, \lambda_{B5} = \lambda_{B6} = .183$
	$\omega_B = .852$	$\theta_{B11} = \theta_{B22} = .019, \theta_{B33} = \theta_{B44} = .027, \theta_{B55} = \theta_{B66} = .034$
.111	$\alpha = .687$	$\alpha = \alpha_w(1 - ICC) + \alpha_B * ICC$
	$\omega = .687$	$\omega = \omega_w(1 - ICC) + \omega_B * ICC$
	$\alpha_w = .667$	$\lambda_{w1} = \lambda_{w2} = .388, \lambda_{w3} = \lambda_{w4} = .399, \lambda_{w5} = \lambda_{w6} = .413$
	$\alpha_B = .850$	$\theta_{w11} = \theta_{w22} = .468, \theta_{w33} = \theta_{w44} = .479, \theta_{w55} = \theta_{w66} = .493$
	$\omega_w = .667$	$\lambda_{B1} = \lambda_{B2} = .137,\ \lambda_{B3} = \lambda_{B4} = .160, \lambda_{B5} = \lambda_{B6} = .183$
	$\omega_B = .852$	$\theta_{B11} = \theta_{B22} = .019, \theta_{B33} = \theta_{B44} = .027, \theta_{B55} = \theta_{B66} = .034$
.296	$\alpha = .514$	$\alpha = \alpha_w(1 - ICC) + \alpha_B * ICC$
	$\omega = .517$	$\omega = \omega_w(1 - ICC) + \omega_B * ICC$
	$\alpha_w = .372$	$\lambda_{w1} = \lambda_{w2} = .121, \lambda_{w3} = \lambda_{w4} = .164, \lambda_{w5} = \lambda_{w6} = .207$
	$\alpha_B = .850$	$\theta_{w11} = \theta_{w22} = .253, \theta_{w33} = \theta_{w44} = .268, \theta_{w55} = \theta_{w66} = .283$
	$\omega_w = .376$	$\lambda_{B1} = \lambda_{B2} = .137,\ \lambda_{B3} = \lambda_{B4} = .160, \lambda_{B5} = \lambda_{B6} = .183$
	$\omega_B = .852$	$\theta_{B11} = \theta_{B22} = .019, \theta_{B33} = \theta_{B44} = .027, \theta_{B55} = \theta_{B66} = .034$
(2) Same Within-group Level Parameters, but Different Between-group Level Parameters		
ICC	Reliability	Parameters
.050	$\alpha = 0.852$	$\alpha = \alpha_w(1 - ICC) + \alpha_B * ICC$
	$\omega = .854$	$\omega = \omega_w(1 - ICC) + \omega_B * ICC$
	$\alpha_w = .852$	$\lambda_{w1} = \lambda_{w2} = .599, \lambda_{w3} = \lambda_{w4} = .699, \lambda_{w5} = \lambda_{w6} = .799$
	$\alpha_B = .850$	$\theta_{w11} = \theta_{w22} = 0.451, \theta_{w33} = \theta_{w44} = 0.501, \theta_{w55} = \theta_{w66} = .551$
	$\omega_w = .854$	$\lambda_{B1} = \lambda_{B2} = .137,\ \lambda_{B3} = \lambda_{B4} = .160, \lambda_{B5} = \lambda_{B6} = .183$
	$\omega_B = .852$	$\theta_{B11} = \theta_{B22} = .019, \theta_{B33} = \theta_{B44} = .027, \theta_{B55} = \theta_{B66} = .034$
.111	$\alpha = .798$	$\alpha = \alpha_w(1 - ICC) + \alpha_B * ICC$
	$\omega = .801$	$\omega = \omega_w(1 - ICC) + \omega_B * ICC$
	$\alpha_w = .852$	$\lambda_{w1} = \lambda_{w2} = .599, \lambda_{w3} = \lambda_{w4} = .699, \lambda_{w5} = \lambda_{w6} = .799$
	$\alpha_B = .372$	$\theta_{w11} = \theta_{w22} = 0.451, \theta_{w33} = \theta_{w44} = 0.501, \theta_{w55} = \theta_{w66} = .551$
	$\omega_w = .854$	$\lambda_{B1} = \lambda_{B2} = .121,\ \lambda_{B3} = \lambda_{B4} = .164, \lambda_{B5} = \lambda_{B6} = .207$
	$\omega_B = .376$	$\theta_{B11} = \theta_{B22} = .253, \theta_{B33} = \theta_{B44} = .268, \theta_{B55} = \theta_{B66} = .283$
.296	$\alpha = .797$	$\alpha = \alpha_w(1 - ICC) + \alpha_B * ICC$
	$\omega = .799$	$\omega = \omega_w(1 - ICC) + \omega_B * ICC$
	$\alpha_w = .852$	$\lambda_{w1} = \lambda_{w2} = .599, \lambda_{w3} = \lambda_{w4} = .699, \lambda_{w5} = \lambda_{w6} = .799$
	$\alpha_B = .667$	$\theta_{w11} = \theta_{w22} = .451, \theta_{w33} = \theta_{w44} = .501, \theta_{w55} = \theta_{w66} = .551$
	$\omega_w = .854$	$\lambda_{B1} = \lambda_{B2} = .388,\ \lambda_{B3} = \lambda_{B4} = .399, \lambda_{B5} = \lambda_{B6} = .413$
	$\omega_B = .667$	$\theta_{B11} = \theta_{B22} = .468, \theta_{B33} = \theta_{B44} = .479, \theta_{B55} = \theta_{B66} = .493$

Note. α=single-level coefficient alpha, α_w=within-group level coefficient alpha, α_B=between-group level coefficient alpha; ω=single-level coefficient omega, ω_w=within-group level coefficient omega, ω_B=between-group level coefficient omega; ICC=intra-class correlation; λ_{wi}=factor loadings at within-group level, λ_{Bi}=factor loadings at between-group level; θ_{wii}=error variances at within-group level, θ_{Bii}=error variances at between-group level.

Table 5

Data Simulation Conditions under MCAR

Cond#	MDM	SS	ICC	MDP	Cond#	MDM	SS	ICC	MDP
1	MCAR	750	.050(B)	0%	37	MCAR	750	.050(W)	0%
2				15%	38				15%
3				30%	39				30%
4			.111(B)	0%	40			.111(W)	0%
5				15%	41				15%
6				30%	42				30%
7			.296(B)	0%	43			.296(W)	0%
8				15%	44				15%
9				30%	45				30%
10		1500(G100)	.050(B)	0%	46		1500(G100)	.050(W)	0%
11				15%	47				15%
12				30%	48				30%
13			.111(B)	0%	49			.111(W)	0%
14				15%	50				15%
15				30%	51				30%
16			.296(B)	0%	52			.296(W)	0%
17				15%	53				15%
18				30%	54				30%
19		1500(G50)	.050(B)	0%	55		1500(G50)	.050(W)	0%
20				15%	56				15%
21				30%	57				30%
22			.111(B)	0%	58			.111(W)	0%
23				15%	59				15%
24				30%	60				30%
25			.296(B)	0%	61			.296(W)	0%
26				15%	62				15%
27				30%	63				30%
28		3000	.050(B)	0%	64		3000	.050(W)	0%
29				15%	65				15%
30				30%	66				30%
31			.111(B)	0%	67			.111(W)	0%
32				15%	68				15%
33				30%	69				30%
34			.296(B)	0%	70			.296(W)	0%
35				15%	71				15%
36				30%	72				30%

Note. Cond#=condition ID; MDM=missing data mechanisms; ICC=intra class correlation, (B)=same between-group level parameters but different within-group level parameters, (W)=same within-group level parameters but different between-group level parameters; SS=sample sizes, (G100)=100 clusters, (G50)=50 clusters; MDP=missing data proportions.

Table 6

Data Simulation Conditions under MAR

Cond#	MDM	SS	ICC	MDP	Cond#	MDM	SS	ICC	MDP
73	MAR	750	.050(B)	0%	109	MAR	750	.050(W)	0%
74				15%	110				15%
75				30%	111				30%
76			.111(B)	0%	112			.111(W)	0%
77				15%	113				15%
78				30%	114				30%
79			.296(B)	0%	115			.296(W)	0%
80				15%	116				15%
81				30%	117				30%
82		1500(G100)	.050(B)	0%	118		1500(G100)	.050(W)	0%
83				15%	119				15%
84				30%	120				30%
85			.111(B)	0%	121			.111(W)	0%
86				15%	122				15%
87				30%	123				30%
88			.296(B)	0%	124			.296(W)	0%
89				15%	125				15%
90				30%	126				30%
91		1500(G50)	.050(B)	0%	127		1500(G50)	.050(W)	0%
92				15%	128				15%
93				30%	129				30%
94			.111(B)	0%	130			.111(W)	0%
95				15%	131				15%
96				30%	132				30%
97			.296(B)	0%	133			.296(W)	0%
98				15%	134				15%
99				30%	135				30%
100		3000	.050(B)	0%	136		3000	.050(W)	0%
101				15%	137				15%
102				30%	138				30%
103			.111(B)	0%	139			.111(W)	0%
104				15%	140				15%
105				30%	141				30%
106			.296(B)	0%	142			.296(W)	0%
107				15%	143				15%
108				30%	144				30%

Note. Cond#=condition ID; MDM=missing data mechanisms; ICC=intra class correlation, (B)=same between-group level parameters but different within-group level parameters, (W)=same within-group level parameters but different between-group level parameters; SS=sample sizes, (G100)=100 clusters, (G50)=50 clusters; MDP=missing data proportions.

Table 7

Data Simulation Conditions under MNAR

Cond#	MDM	SS	ICC	MDP	Cond#	MDM	SS	ICC	MDP
145	MNAR	750	.050(B)	0%	181	MNAR	750	.050(W)	0%
146				15%	182				15%
147				30%	183				30%
148			.111(B)	0%	184			.111(W)	0%
149				15%	185				15%
150				30%	186				30%
151			.296(B)	0%	187			.296(W)	0%
152				15%	188				15%
153				30%	189				30%
154		1500(G100)	.050(B)	0%	190		1500(G100)	.050(W)	0%
155				15%	191				15%
156				30%	192				30%
157			.111(B)	0%	193			.111(W)	0%
158				15%	194				15%
159				30%	195				30%
160			.296(B)	0%	196			.296(W)	0%
161				15%	197				15%
162				30%	198				30%
163		1500(G50)	.050(B)	0%	199		1500(G50)	.050(W)	0%
164				15%	200				15%
165				30%	201				30%
166			.111(B)	0%	202			.111(W)	0%
167				15%	203				15%
168				30%	204				30%
169			.296(B)	0%	205			.296(W)	0%
170				15%	206				15%
171				30%	207				30%
172		3000	.050(B)	0%	208		3000	.050(W)	0%
173				15%	209				15%
174				30%	210				30%
175			.111(B)	0%	211			.111(W)	0%
176				15%	212				15%
177				30%	213				30%
178			.296(B)	0%	214			.296(W)	0%
179				15%	215				15%
180				30%	216				30%

Note. Cond#=condition ID; MDM=missing data mechanisms; ICC=intra class correlation, (B)=same between-group level parameters but different within-group level parameters, (W)=same within-group level parameters but different between-group level parameters; SS=sample sizes, (G100)=100 clusters, (G50)=50 clusters; MDP=missing data proportions.

Table 8

Results of α under MCAR using LD for Same Between-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
1	MCAR	750	.050(B)	0	100.0	1.5	100.0	1.5	100.0	-1.5
2				15	100.0	1.5	100.0	1.5	100.0	-3.1
3				30	100.0	1.4	100.0	1.4	100.0	-4.7
4			.111(B)	0	100.0	2.7	100.0	2.8	100.0	-0.2
5				15	100.0	2.5	100.0	2.7	100.0	-2.5
6				30	100.0	2.6	100.0	3.0	100.0	-7.1
7			.296(B)	0	100.0	3.4	100.0	5.3	100.0	1.2
8				15	100.0	3.3	100.0	5.5	100.0	0.8
9				30	100.0	2.9	100.0	5.7	100.0	-1.5
10		1500(G100)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
11				15	100.0	1.5	100.0	1.5	100.0	-1.0
12				30	100.0	1.5	100.0	1.4	100.0	-2.5
13			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.4
14				15	100.0	2.6	100.0	2.6	100.0	0.2
15				30	100.0	2.6	100.0	2.8	100.0	-2.5
16			.296(B)	0	100.0	3.7	100.0	5.3	100.0	1.7
17				15	100.0	3.8	100.0	5.2	100.0	1.8
18				30	100.0	3.8	100.0	5.8	100.0	1.0
19		1500(G50)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
20				15	100.0	1.5	100.0	1.5	100.0	-1.0
21				30	100.0	1.5	100.0	1.5	100.0	-2.5
22			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.2
23				15	100.0	2.6	100.0	2.7	100.0	0.5
24				30	100.0	2.5	100.0	2.6	100.0	-1.4
25			.296(B)	0	100.0	3.6	100.0	5.2	100.0	1.5
26				15	100.0	3.5	100.0	4.9	100.0	1.5
27				30	100.0	3.3	100.0	4.9	100.0	0.8
28		3000	.050(B)	0	100.0	1.6	100.0	1.5	100.0	1.2
29				15	100.0	1.5	100.0	1.5	100.0	0.8
30				30	100.0	1.6	100.0	1.6	100.0	-1.1
31			.111(B)	0	100.0	2.8	100.0	2.8	100.0	1.7
32				15	100.0	2.7	100.0	2.8	100.0	1.4
33				30	100.0	2.7	100.0	2.8	100.0	0.8
34			.296(B)	0	100.0	4.0	100.0	5.6	100.0	1.8
35				15	100.0	4.0	100.0	5.7	100.0	1.8
36				30	100.0	4.1	100.0	5.7	100.0	1.8

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 9

Results of α under MCAR using LD for Same Within-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
37	MCAR	750	.050$^{(W)}$	0	100.0	1.5	100.0	1.5	100.0	-1.5
38				15	100.0	1.5	100.0	1.4	100.0	-2.6
39				30	100.0	1.4	100.0	1.4	100.0	-5.2
40			.111$^{(W)}$	0	100.0	2.0	100.0	1.5	100.0	5.8
41				15	100.0	1.9	100.0	1.4	100.0	9.5
42				30	100.0	1.9	100.0	1.2	100.0	19.2
43			.296$^{(W)}$	0	100.0	2.3	100.0	2.0	100.0	-0.5
44				15	100.0	2.2	100.0	2.0	100.0	-1.1
45				30	100.0	2.1	100.0	1.8	100.0	-2.5
46		1500$^{(G100)}$	.050$^{(W)}$	0	100.0	1.5	100.0	1.5	100.0	0.1
47				15	100.0	1.5	100.0	1.5	100.0	-1.6
48				30	100.0	1.4	100.0	1.4	100.0	-4.3
49			.111$^{(W)}$	0	100.0	1.9	100.0	1.5	100.0	0.7
50				15	100.0	1.9	100.0	1.5	100.0	2.2
51				30	100.0	1.8	100.0	1.3	100.0	6.5
52			.296$^{(W)}$	0	100.0	2.3	100.0	2.0	100.0	1.2
53				15	100.0	2.2	100.0	1.9	100.0	1.0
54				30	100.0	2.3	100.0	2.0	100.0	-0.3
55		1500$^{(G50)}$	.050$^{(W)}$	0	100.0	1.5	100.0	1.5	100.0	0.1
56				15	100.0	1.5	100.0	1.5	100.0	-1.0
57				30	100.0	1.5	100.0	1.5	100.0	-2.6
58			.111$^{(W)}$	0	100.0	2.0	100.0	1.5	100.0	3.1
59				15	100.0	2.0	100.0	1.5	100.0	5.6
60				30	100.0	2.0	100.0	1.5	100.0	7.1
61			.296$^{(W)}$	0	100.0	2.3	100.0	2.0	100.0	0.4
62				15	100.0	2.3	100.0	2.0	100.0	0.1
63				30	100.0	2.2	100.0	1.9	100.0	-0.7
64		3000	.050$^{(W)}$	0	100.0	1.6	100.0	1.5	100.0	1.2
65				15	100.0	1.6	100.0	1.6	100.0	0.5
66				30	100.0	1.6	100.0	1.5	100.0	-1.3
67			.111$^{(W)}$	0	100.0	2.0	100.0	1.5	100.0	1.7
68				15	100.0	1.9	100.0	1.5	100.0	0.8
69				30	100.0	2.0	100.0	1.5	100.0	1.9
70			.296$^{(W)}$	0	100.0	2.4	100.0	2.1	100.0	1.7
71				15	100.0	2.3	100.0	2.0	100.0	1.5
72				30	100.0	2.3	100.0	2.0	100.0	1.2

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 10

Results of ω under MCAR using LD for Same Between-Group Level Parameters Conditions

					ω		ω_w		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
1	MCAR	750	.050(B)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
2				15	100.0	-0.1	96.6	0.0	96.6	-4.4
3				30	99.1	-0.1	94.2	0.0	94.2	-3.6
4			.111(B)	0	100.0	-0.1	98.9	0.0	98.9	-0.4
5				15	99.8	-0.2	93.2	0.0	93.2	-1.9
6				30	97.2	0.1	77.5	0.8	77.5	-3.5
7			.296(B)	0	96.2	-0.6	82.6	0.6	82.6	0.3
8				15	91.0	-0.2	67.3	1.7	67.3	0.6
9				30	80.4	0.0	52.9	3.6	52.9	0.2
10		1500(G100)	.050(B)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
11				15	100.0	-0.1	99.7	0.0	99.7	-2.8
12				30	100.0	-0.1	98.6	0.0	98.6	-3.1
13			.111(B)	0	100.0	-0.1	100.0	-0.1	100.0	0.0
14				15	100.0	-0.2	99.6	-0.2	99.6	-0.3
15				30	100.0	-0.1	94.9	0.1	94.9	-1.4
16			.296(B)	0	99.7	-0.6	97.9	-0.1	97.9	0.1
17				15	98.4	-0.3	91.6	0.2	91.6	0.4
18				30	95.4	0.0	74.8	1.5	74.8	0.5
19		1500(G50)	.050(B)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
20				15	100.0	-0.1	99.4	0.0	99.4	-2.2
21				30	100.0	0.0	98.8	0.0	98.8	-3.8
22			.111(B)	0	100.0	-0.1	99.9	-0.2	99.9	0.0
23				15	100.0	-0.2	99.0	-0.2	99.0	-0.2
24				30	100.0	-0.2	95.7	-0.1	95.7	-1.0
25			.296(B)	0	99.5	-0.6	94.1	-0.1	94.1	0.1
26				15	98.0	-0.5	87.5	0.1	87.5	0.4
27				30	93.5	-0.3	69.8	1.3	69.8	0.4
28		3000	.050(B)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
29				15	100.0	-0.1	99.7	0.0	99.7	-0.5
30				30	100.0	0.0	99.1	0.1	99.1	-3.0
31			.111(B)	0	100.0	0.0	100.0	-0.1	100.0	0.0
32				15	100.0	0.0	99.9	-0.1	99.9	-0.1
33				30	100.0	0.0	99.5	0.0	99.5	-0.1
34			.296(B)	0	100.0	-0.4	99.8	-0.1	99.8	0.0
35				15	100.0	-0.3	98.8	0.2	98.8	0.1
36				30	98.9	-0.1	94.5	0.6	94.5	0.4

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_w=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 11

Results of ω under MCAR using LD for Same Within-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
37	MCAR	750	.050(W)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
38				15	100.0	-0.1	98.1	0.0	98.1	-3.7
39				30	99.5	-0.1	91.8	0.0	91.8	-3.8
40			.111(W)	0	97.0	0.0	100.0	0.0	100.0	5.8
41				15	95.0	0.0	98.1	0.0	98.1	6.6
42				30	89.2	0.0	95.9	-0.1	95.9	12.1
43			.296(W)	0	77.8	0.0	99.7	0.0	99.7	-1.8
44				15	76.3	0.0	99.3	0.0	99.3	-2.7
45				30	72.2	-0.1	97.9	-0.1	97.9	-4.2
46		1500(G100)	.050(W)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
47				15	100.0	-0.1	99.6	0.0	99.6	-3.2
48				30	100.0	-0.2	98.4	0.0	98.4	-4.9
49			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	1.1
50				15	99.9	-0.1	98.9	0.0	98.9	1.8
51				30	99.8	-0.2	96.2	-0.1	96.2	5.8
52			.296(W)	0	99.2	-0.1	100.0	0.0	100.0	-0.6
53				15	98.2	-0.1	99.9	-0.1	99.9	-0.8
54				30	96.5	0.0	100.0	0.0	100.0	-2.2
55		1500(G50)	.050(W)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
56				15	100.0	-0.1	99.1	0.0	99.1	-3.1
57				30	100.0	-0.1	97.7	0.0	97.7	-3.7
58			.111(W)	0	98.4	0.0	99.9	0.0	99.9	6.3
59				15	97.8	0.0	99.8	0.0	99.8	5.8
60				30	96.0	0.1	98.4	0.0	98.4	7.1
61			.296(W)	0	79.4	0.1	99.7	0.0	99.7	-0.9
62				15	79.1	0.2	99.8	0.0	99.8	-1.2
63				30	75.1	0.0	99.8	-0.1	99.8	-2.0
64		3000	.050(W)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
65				15	100.0	0.0	100.0	0.0	100.0	-0.7
66				30	100.0	0.0	99.7	0.0	99.7	-3.3
67			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	2.5
68				15	99.9	-0.1	100.0	0.0	100.0	1.4
69				30	100.0	0.0	99.4	0.0	99.4	2.0
70			.296(W)	0	99.4	0.0	99.8	0.0	99.8	-0.4
71				15	99.1	-0.1	100.0	0.0	100.0	-0.5
72				30	98.3	0.0	99.9	0.0	99.9	-0.7

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 12

Results of α under MCAR using FIML for Same Between-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
1	MCAR	750	.050(B)	0	100.0	1.5	100.0	1.5	100.0	-1.5
2				15	100.0	1.5	100.0	1.5	100.0	-2.2
3				30	100.0	1.5	100.0	1.5	100.0	-2.4
4			.111(B)	0	100.0	2.7	100.0	2.8	100.0	-0.2
5				15	100.0	2.6	100.0	2.7	100.0	-0.5
6				30	100.0	2.7	100.0	2.9	100.0	-0.6
7			.296(B)	0	100.0	3.4	100.0	5.3	100.0	1.2
8				15	100.0	3.4	100.0	5.4	100.0	1.2
9				30	100.0	3.3	100.0	5.0	100.0	1.3
10		1500(G100)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
11				15	100.0	1.5	100.0	1.5	100.0	-0.1
12				30	100.0	1.6	100.0	1.5	100.0	0.0
13			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.4
14				15	100.0	2.7	100.0	2.7	100.0	1.3
15				30	100.0	2.7	100.0	2.7	100.0	1.3
16			.296(B)	0	100.0	3.7	100.0	5.3	100.0	1.7
17				15	100.0	3.8	100.0	5.4	100.0	1.8
18				30	100.0	3.7	100.0	5.4	100.0	1.7
19		1500(G50)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
20				15	100.0	1.5	100.0	1.5	100.0	0.1
21				30	100.0	1.5	100.0	1.5	100.0	0.0
22			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.2
23				15	100.0	2.7	100.0	2.6	100.0	1.2
24				30	100.0	2.7	100.0	2.7	100.0	1.2
25			.296(B)	0	100.0	3.6	100.0	5.2	100.0	1.5
26				15	100.0	3.6	100.0	5.1	100.0	1.5
27				30	100.0	3.6	100.0	5.3	100.0	1.4
28		3000	.050(B)	0	100.0	1.6	100.0	1.5	100.0	1.2
29				15	100.0	1.6	100.0	1.5	100.0	1.2
30				30	100.0	1.6	100.0	1.5	100.0	1.3
31			.111(B)	0	100.0	2.8	100.0	2.8	100.0	1.7
32				15	100.0	2.8	100.0	2.8	100.0	1.7
33				30	100.0	2.8	100.0	2.8	100.0	1.7
34			.296(B)	0	100.0	4.0	100.0	5.6	100.0	1.8
35				15	100.0	4.0	100.0	5.6	100.0	1.8
36				30	100.0	4.1	100.0	5.8	100.0	1.8

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 13

Results of α under MCAR using FIML for Same Within-Group Level Parameters Conditions

					α		α_W		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
37	MCAR	750	.050(W)	0	100.0	1.5	100.0	1.5	100.0	-1.5
38				15	100.0	1.5	100.0	1.5	100.0	-1.9
39				30	100.0	1.5	100.0	1.5	100.0	-2.4
40			.111(W)	0	100.0	2.0	100.0	1.5	100.0	5.8
41				15	100.0	2.0	100.0	1.5	100.0	5.3
42				30	100.0	1.9	100.0	1.5	100.0	6.3
43			.296(W)	0	100.0	2.3	100.0	2.0	100.0	-0.5
44				15	100.0	2.3	100.0	2.0	100.0	-0.6
45				30	100.0	2.3	100.0	2.0	100.0	-0.6
46		1500(G100)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
47				15	100.0	1.5	100.0	1.5	100.0	-0.1
48				30	100.0	1.5	100.0	1.5	100.0	-0.1
49			.111(W)	0	100.0	1.9	100.0	1.5	100.0	0.7
50				15	100.0	1.9	100.0	1.5	100.0	0.6
51				30	100.0	1.9	100.0	1.5	100.0	0.5
52			.296(W)	0	100.0	2.3	100.0	2.0	100.0	1.2
53				15	100.0	2.3	100.0	2.0	100.0	1.2
54				30	100.0	2.3	100.0	2.0	100.0	1.2
55		1500(G50)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
56				15	100.0	1.5	100.0	1.5	100.0	0.1
57				30	100.0	1.5	100.0	1.5	100.0	0.0
58			.111(W)	0	100.0	2.0	100.0	1.5	100.0	3.1
59				15	100.0	2.0	100.0	1.5	100.0	3.4
60				30	100.0	2.0	100.0	1.5	100.0	3.5
61			.296(W)	0	100.0	2.3	100.0	2.0	100.0	0.4
62				15	100.0	2.3	100.0	2.0	100.0	0.4
63				30	100.0	2.3	100.0	2.0	100.0	0.4
64		3000	.050(W)	0	100.0	1.6	100.0	1.5	100.0	1.2
65				15	100.0	1.6	100.0	1.5	100.0	1.2
66				30	100.0	1.6	100.0	1.5	100.0	1.2
67			.111(W)	0	100.0	2.0	100.0	1.5	100.0	1.7
68				15	100.0	2.0	100.0	1.6	100.0	1.7
69				30	100.0	2.0	100.0	1.5	100.0	1.6
70			.296(W)	0	100.0	2.4	100.0	2.1	100.0	1.7
71				15	100.0	2.4	100.0	2.1	100.0	1.6
72				30	100.0	2.4	100.0	2.1	100.0	1.7

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 14

Results of ω under MCAR using FIML for Same Between-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
1	MCAR	750	.050(B)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
2				15	100.0	-0.1	99.3	0.0	99.3	-3.6
3				30	100.0	-0.1	98.9	0.0	98.9	-3.9
4			.111(B)	0	100.0	-0.1	98.9	0.0	98.9	-0.4
5				15	100.0	-0.1	98.5	0.0	98.5	-0.5
6				30	100.0	0.0	97.5	0.1	97.5	-0.6
7			.296(B)	0	96.2	-0.6	82.6	0.6	82.6	0.3
8				15	94.9	-0.4	77.1	1.2	77.1	0.4
9				30	91.6	-0.3	74.8	0.8	74.8	0.4
10		1500(G100)	.050(B)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
11				15	100.0	-0.1	100.0	0.0	100.0	-1.6
12				30	100.0	-0.1	99.4	0.0	99.4	-1.2
13			.111(B)	0	100.0	-0.1	100.0	-0.1	100.0	0.0
14				15	100.0	-0.1	99.8	-0.1	99.8	0.1
15				30	100.0	0.0	99.8	-0.1	99.8	0.1
16			.296(B)	0	99.7	-0.6	97.9	-0.1	97.9	0.1
17				15	99.7	-0.4	96.1	0.1	96.1	0.2
18				30	98.9	-0.4	95.5	0.1	95.5	0.2
19		1500(G50)	.050(B)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
20				15	100.0	-0.1	100.0	0.0	100.0	-1.2
21				30	100.0	-0.1	99.5	0.0	99.5	-1.0
22			.111(B)	0	100.0	-0.1	99.9	-0.2	99.9	0.0
23				15	100.0	-0.1	100.0	-0.2	100.0	0.1
24				30	100.0	-0.1	99.9	-0.2	99.9	0.2
25			.296(B)	0	99.5	-0.6	94.1	-0.1	94.1	0.1
26				15	99.5	-0.6	92.1	-0.1	92.1	0.2
27				30	98.6	-0.5	89.7	0.4	89.7	0.2
28		3000	.050(B)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
29				15	100.0	-0.1	100.0	0.0	100.0	-0.3
30				30	100.0	-0.1	99.9	0.0	99.9	-0.2
31			.111(B)	0	100.0	0.0	100.0	-0.1	100.0	0.0
32				15	100.0	0.0	100.0	-0.1	100.0	0.0
33				30	100.0	0.0	100.0	-0.1	100.0	0.1
34			.296(B)	0	100.0	-0.4	99.8	-0.1	99.8	0.0
35				15	100.0	-0.4	99.9	0.0	99.9	0.0
36				30	100.0	-0.3	99.7	0.1	99.7	0.1

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 15

Results of ω under MCAR using FIML for Same Within-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
37	MCAR	750	.050(W)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
38				15	100.0	-0.1	99.0	0.0	99.0	-3.7
39				30	100.0	-0.1	99.0	0.0	99.0	-3.4
40			.111(W)	0	100.0	0.0	100.0	0.0	100.0	5.8
41				15	99.1	0.0	99.8	0.0	99.8	5.6
42				30	97.0	0.0	99.5	0.0	99.5	6.6
43			.296(W)	0	93.5	0.0	99.7	0.0	99.7	-1.8
44				15	87.3	0.0	99.6	0.0	99.6	-1.8
45				30	77.8	0.0	99.7	0.0	99.7	-2.0
46		1500(G100)	.050(W)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
47				15	100.0	-0.1	99.6	0.0	99.6	-1.6
48				30	100.0	-0.1	99.7	0.0	99.7	-1.3
49			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	1.1
50				15	100.0	-0.1	99.8	0.0	99.8	1.2
51				30	100.0	-0.1	99.9	0.0	99.9	1.5
52			.296(W)	0	99.9	-0.1	100.0	0.0	100.0	-0.6
53				15	99.5	-0.1	99.8	0.0	99.8	-0.6
54				30	99.2	0.0	99.9	0.0	99.9	-0.7
55		1500(G50)	.050(W)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
56				15	100.0	-0.1	99.6	0.0	99.6	-1.2
57				30	100.0	-0.1	99.9	0.0	99.9	-1.2
58			.111(W)	0	99.9	0.0	99.9	0.0	99.9	6.3
59				15	99.6	0.0	99.9	0.0	99.9	6.4
60				30	98.4	0.0	99.8	0.0	99.8	6.1
61			.296(W)	0	94.3	0.1	100.0	0.0	100.0	-0.9
62				15	88.5	0.1	99.9	0.0	99.9	-0.9
63				30	79.4	0.0	99.7	0.0	99.7	-0.9
64		3000	.050(W)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
65				15	100.0	-0.1	100.0	0.0	100.0	-0.3
66				30	100.0	-0.1	100.0	0.0	100.0	-0.3
67			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	2.5
68				15	100.0	-0.1	100.0	0.0	100.0	2.4
69				30	100.0	-0.1	99.9	0.0	99.9	2.5
70			.296(W)	0	99.9	0.0	100.0	0.0	100.0	-0.4
71				15	99.7	0.0	100.0	0.0	100.0	-0.4
72				30	99.4	0.0	99.8	0.0	99.8	-0.3

Note. #=condition ID regarding Table 5; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 16

Results of α under MAR using LD for Same Between-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
73	MAR	750	.050(B)	0	100.0	1.5	100.0	1.5	100.0	-1.5
74				15	100.0	-6.3	100.0	-6.2	100.0	-14.7
75				30	100.0	-10.6	100.0	-10.6	100.0	-20.7
76			.111(B)	0	100.0	2.7	100.0	2.8	100.0	-0.2
77				15	100.0	-9.2	100.0	-8.6	100.0	-13.3
78				30	100.0	-15.2	100.0	-14.6	100.0	-22.7
79			.296(B)	0	100.0	3.4	100.0	5.3	100.0	1.2
80				15	100.0	-12.6	100.0	-13.5	100.0	-5.2
81				30	100.0	-20.1	100.0	-21.5	100.0	-12.2
82		1500(G100)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
83				15	100.0	-6.3	100.0	-6.2	100.0	-14.8
84				30	100.0	-10.6	100.0	-10.4	100.0	-20.3
85			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.4
86				15	100.0	-9.2	100.0	-8.7	100.0	-10.8
87				30	100.0	-14.9	100.0	-14.2	100.0	-20.0
88			.296(B)	0	100.0	3.7	100.0	5.3	100.0	1.7
89				15	100.0	-12.4	100.0	-14.0	100.0	-3.7
90				30	100.0	-19.5	100.0	-21.4	100.0	-9.1
91		1500(G50)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
92				15	100.0	-6.3	100.0	-6.2	100.0	-15.4
93				30	100.0	-10.6	100.0	-10.5	100.0	-20.0
94			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.2
95				15	100.0	-9.3	100.0	-8.9	100.0	-10.3
96				30	100.0	-15.0	100.0	-14.5	100.0	-18.6
97			.296(B)	0	100.0	3.6	100.0	5.2	100.0	1.5
98				15	100.0	-12.5	100.0	-14.4	100.0	-3.7
99				30	100.0	-19.7	100.0	-22.4	100.0	-8.4
100		3000	.050(B)	0	100.0	1.6	100.0	1.5	100.0	1.2
101				15	100.0	-6.2	100.0	-6.1	100.0	-14.2
102				30	100.0	-10.5	100.0	-10.3	100.0	-20.2
103			.111(B)	0	100.0	2.8	100.0	2.8	100.0	1.7
104				15	100.0	-9.0	100.0	-8.7	100.0	-8.2
105				30	100.0	-14.6	100.0	-14.2	100.0	-15.3
106			.296(B)	0	100.0	4.0	100.0	5.6	100.0	1.8
107				15	100.0	-11.8	100.0	-13.8	100.0	-3.1
108				30	100.0	-18.7	100.0	-21.6	100.0	-6.9

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 17

Results of α under MAR using LD for Same Within-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
109	MAR	750	.050(W)	0	100.0	1.5	100.0	1.5	100.0	-1.5
110				15	100.0	-6.3	100.0	-6.2	100.0	-14.7
111				30	100.0	-10.6	100.0	-10.6	100.0	-20.7
112			.111(W)	0	100.0	2.0	100.0	1.5	100.0	5.8
113				15	100.0	-7.9	100.0	-5.7	100.0	-28.8
114				30	100.0	-13.0	100.0	-10.1	100.0	-25.8
115			.296(W)	0	100.0	2.3	100.0	2.0	100.0	-0.5
116				15	100.0	-7.0	100.0	-3.3	100.0	-37.8
117				30	100.0	-11.8	100.0	-6.6	100.0	-46.7
118		1500(G100)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
119				15	100.0	-6.3	100.0	-6.2	100.0	-14.8
120				30	100.0	-10.6	100.0	-10.4	100.0	-20.3
121			.111(W)	0	100.0	1.9	100.0	1.5	100.0	0.7
122				15	100.0	-8.0	100.0	-5.6	100.0	-43.3
123				30	100.0	-13.1	100.0	-9.8	100.0	-41.1
124			.296(W)	0	100.0	2.3	100.0	2.0	100.0	1.2
125				15	100.0	-7.0	100.0	-3.3	100.0	-38.2
126				30	100.0	-11.6	100.0	-6.4	100.0	-50.1
127		1500(G50)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
128				15	100.0	-6.3	100.0	-6.2	100.0	-15.4
129				30	100.0	-10.6	100.0	-10.5	100.0	-20.0
130			.111(W)	0	100.0	2.0	100.0	1.5	100.0	3.1
131				15	100.0	-7.9	100.0	-5.6	100.0	-37.2
132				30	100.0	-12.9	100.0	-9.7	100.0	-37.5
133			.296(W)	0	100.0	2.3	100.0	2.0	100.0	0.4
134				15	100.0	-6.9	100.0	-3.3	100.0	-37.0
135				30	100.0	-11.5	100.0	-6.4	100.0	-49.1
136		3000	.050(W)	0	100.0	1.6	100.0	1.5	100.0	1.2
137				15	100.0	-6.2	100.0	-6.1	100.0	-14.2
138				30	100.0	-10.5	100.0	-10.3	100.0	-20.2
139			.111(W)	0	100.0	2.0	100.0	1.5	100.0	1.7
140				15	100.0	-7.9	100.0	-5.5	100.0	-51.3
141				30	100.0	-12.9	100.0	-9.5	100.0	-50.5
142			.296(W)	0	100.0	2.4	100.0	2.1	100.0	1.7
143				15	100.0	-6.8	100.0	-3.2	100.0	-36.6
144				30	100.0	-11.4	100.0	-6.3	100.0	-51.6

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 18

Results of ω under MAR using LD for Same Between-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
73	MAR	750	.050(B)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
74				15	100.0	-8.5	97.8	-8.3	97.8	-15.1
75				30	100.0	-13.0	92.9	-12.7	92.9	-17.3
76			.111(B)	0	100.0	-0.1	98.9	0.0	98.9	-0.4
77				15	98.3	-12.2	83.4	-11.4	83.4	-11.7
78				30	90.5	-17.7	63.9	-16.5	63.9	-16.7
79			.296(B)	0	96.2	-0.6	82.6	0.6	82.6	0.3
80				15	80.9	-16.0	55.8	-16.2	55.8	-3.9
81				30	70.0	-22.5	39.8	-20.8	39.8	-7.4
82		1500(G100)	.050(B)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
83				15	100.0	-8.5	99.0	-8.3	99.0	-15.4
84				30	100.0	-13.1	97.1	-12.8	97.1	-19.6
85			.111(B)	0	100.0	-0.1	100.0	-0.1	100.0	0.0
86				15	100.0	-12.4	97.8	-11.9	97.8	-10.2
87				30	98.4	-17.9	86.5	-17.0	86.5	-17.6
88			.296(B)	0	99.7	-0.6	97.9	-0.1	97.9	0.1
89				15	92.8	-16.4	79.1	-17.4	79.1	-3.9
90				30	79.9	-22.7	56.3	-23.9	56.3	-7.0
91		1500(G50)	.050(B)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
92				15	100.0	-8.5	99.2	-8.3	99.2	-15.2
93				30	100.0	-13.0	98.3	-12.8	98.3	-18.8
94			.111(B)	0	100.0	-0.1	99.9	-0.2	99.9	0.0
95				15	99.9	-12.4	97.1	-12.1	97.1	-9.7
96				30	97.7	-17.9	83.8	-17.3	83.8	-14.2
97			.296(B)	0	99.5	-0.6	94.1	-0.1	94.1	0.1
98				15	90.2	-16.3	74.6	-17.6	74.6	-3.6
99				30	80.8	-23.0	50.4	-24.7	50.4	-5.6
100		3000	.050(B)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
101				15	100.0	-8.5	99.6	-8.2	99.6	-13.0
102				30	100.0	-13.0	99.5	-12.7	99.5	-18.1
103			.111(B)	0	100.0	0.0	100.0	-0.1	100.0	0.0
104				15	100.0	-12.2	100.0	-12.0	100.0	-8.9
105				30	99.9	-17.8	97.9	-17.4	97.9	-13.7
106			.296(B)	0	100.0	-0.4	99.8	-0.1	99.8	0.0
107				15	98.6	-16.3	94.8	-18.0	94.8	-3.9
108				30	93.0	-22.8	78.1	-25.6	78.1	-6.5

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 19

Results of ω under MAR using LD for Same Within-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
109	MAR	750	.050(W)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
110				15	100.0	-8.5	97.8	-8.3	97.8	-15.1
111				30	100.0	-13.0	92.8	-12.7	92.8	-17.3
112			.111(W)	0	97.0	0.0	100.0	0.0	100.0	5.8
113				15	96.4	-10.5	98.3	-7.8	98.3	-20.7
114				30	87.6	-15.6	95.0	-12.3	95.0	-17.0
115			.296(W)	0	77.8	0.0	99.7	0.0	99.7	-1.8
116				15	77.0	-9.2	99.8	-5.4	99.8	-36.5
117				30	63.4	-13.3	97.8	-8.6	97.8	-44.0
118		1500(G100)	.050(W)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
119				15	100.0	-8.5	99.0	-8.3	99.0	-15.4
120				30	100.0	-13.1	97.1	-12.8	97.1	-19.6
121			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	1.1
122				15	100.0	-10.7	99.2	-7.8	99.2	-29.2
123				30	98.2	-15.9	97.8	-12.3	97.8	-25.8
124			.296(W)	0	99.2	-0.1	100.0	0.0	100.0	-0.6
125				15	98.1	-9.5	99.7	-5.4	99.7	-37.3
126				30	90.4	-13.8	99.4	-8.5	99.4	-45.7
127		1500(G50)	.050(W)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
128				15	100.0	-8.5	99.2	-8.3	99.2	-15.2
129				30	100.0	-13.0	98.3	-12.8	98.3	-18.8
130			.111(W)	0	98.4	0.0	99.9	0.0	99.9	6.3
131				15	97.7	-10.5	99.6	-7.8	99.6	-26.3
132				30	93.3	-15.7	99.1	-12.1	99.1	-23.5
133			.296(W)	0	79.4	0.1	99.7	0.0	99.7	-0.9
134				15	78.7	-9.1	99.7	-5.5	99.7	-34.6
135				30	66.3	-13.3	99.9	-8.6	99.9	-43.5
136		3000	.050(W)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
137				15	100.0	-8.5	99.6	-8.2	99.6	-13.0
138				30	100.0	-13.0	99.5	-12.7	99.5	-18.1
139			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	2.5
140				15	100.0	-10.6	100.0	-7.7	100.0	-34.1
141				30	100.0	-15.8	99.7	-12.0	99.7	-34.2
142			.296(W)	0	99.4	0.0	99.8	0.0	99.8	-0.4
143				15	99.3	-9.3	99.9	-5.4	99.9	-35.7
144				30	94.8	-13.8	99.9	-8.5	99.9	-45.3

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 20

Results of α under MAR using FIML for Same Between-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
73	MAR	750	.050(B)	0	100.0	1.5	100.0	1.5	100.0	-1.5
74				15	100.0	1.5	100.0	1.5	100.0	-1.9
75				30	100.0	1.5	100.0	1.5	100.0	-2.3
76			.111(B)	0	100.0	2.7	100.0	2.8	100.0	-0.2
77				15	100.0	2.7	100.0	2.8	100.0	-0.4
78				30	100.0	2.6	100.0	2.8	100.0	-0.9
79			.296(B)	0	100.0	3.4	100.0	5.3	100.0	1.2
80				15	100.0	2.4	100.0	3.0	100.0	1.2
81				30	100.0	2.2	100.0	2.8	100.0	1.1
82		1500(G100)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
83				15	100.0	1.5	100.0	1.5	100.0	-0.1
84				30	100.0	1.5	100.0	1.5	100.0	-0.2
85			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.4
86				15	100.0	2.7	100.0	2.8	100.0	1.3
87				30	100.0	2.7	100.0	2.7	100.0	1.2
88			.296(B)	0	100.0	3.7	100.0	5.3	100.0	1.7
89				15	100.0	2.7	100.0	3.0	100.0	1.7
90				30	100.0	2.6	100.0	2.8	100.0	1.7
91		1500(G50)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
92				15	100.0	1.5	100.0	1.5	100.0	0.0
93				30	100.0	1.5	100.0	1.5	100.0	0.1
94			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.2
95				15	100.0	2.7	100.0	2.7	100.0	1.1
96				30	100.0	2.7	100.0	2.7	100.0	1.2
97			.296(B)	0	100.0	3.6	100.0	5.2	100.0	1.5
98				15	100.0	2.5	100.0	2.8	100.0	1.4
99				30	100.0	2.6	100.0	2.9	100.0	1.5
100		3000	.050(B)	0	100.0	1.6	100.0	1.5	100.0	1.2
101				15	100.0	1.6	100.0	1.5	100.0	1.1
102				30	100.0	1.6	100.0	1.5	100.0	1.1
103			.111(B)	0	100.0	2.8	100.0	2.8	100.0	1.7
104				15	100.0	2.8	100.0	2.8	100.0	1.7
105				30	100.0	2.8	100.0	2.8	100.0	1.7
106			.296(B)	0	100.0	4.0	100.0	5.6	100.0	1.8
107				15	100.0	3.0	100.0	3.2	100.0	1.8
108				30	100.0	3.0	100.0	3.2	100.0	1.8

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 21

Results of α under MAR using FIML for Same Within-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
109	MAR	750	.050(W)	0	100.0	1.5	100.0	1.5	100.0	-1.5
110				15	100.0	1.5	100.0	1.5	100.0	-1.9
111				30	100.0	1.5	100.0	1.5	100.0	-2.3
112			.111(W)	0	100.0	2.0	100.0	1.5	100.0	5.8
113				15	100.0	1.9	100.0	1.5	100.0	4.9
114				30	100.0	1.9	100.0	1.5	100.0	5.3
115			.296(W)	0	100.0	2.3	100.0	2.0	100.0	-0.5
116				15	100.0	2.2	100.0	2.0	100.0	-0.6
117				30	100.0	2.2	100.0	2.0	100.0	-0.8
118		1500(G100)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
119				15	100.0	1.5	100.0	1.5	100.0	-0.1
120				30	100.0	1.5	100.0	1.5	100.0	-0.2
121			.111(W)	0	100.0	1.9	100.0	1.5	100.0	0.7
122				15	100.0	1.9	100.0	1.5	100.0	0.7
123				30	100.0	1.9	100.0	1.5	100.0	0.5
124			.296(W)	0	100.0	2.3	100.0	2.0	100.0	1.2
125				15	100.0	2.3	100.0	2.0	100.0	1.1
126				30	100.0	2.3	100.0	2.0	100.0	1.0
127		1500(G50)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
128				15	100.0	1.5	100.0	1.5	100.0	0.0
129				30	100.0	1.5	100.0	1.5	100.0	0.1
130			.111(W)	0	100.0	2.0	100.0	1.5	100.0	3.1
131				15	100.0	2.0	100.0	1.5	100.0	3.5
132				30	100.0	2.0	100.0	1.5	100.0	3.8
133			.296(W)	0	100.0	2.3	100.0	2.0	100.0	0.4
134				15	100.0	2.3	100.0	2.0	100.0	0.3
135				30	100.0	2.3	100.0	2.0	100.0	0.3
136		3000	.050(W)	0	100.0	1.6	100.0	1.5	100.0	1.2
137				15	100.0	1.6	100.0	1.5	100.0	1.1
138				30	100.0	1.6	100.0	1.5	100.0	1.1
139			.111(W)	0	100.0	2.0	100.0	1.5	100.0	1.7
140				15	100.0	1.9	100.0	1.5	100.0	1.0
141				30	100.0	2.0	100.0	1.6	100.0	1.6
142			.296(W)	0	100.0	2.4	100.0	2.1	100.0	1.7
143				15	100.0	2.3	100.0	2.1	100.0	1.7
144				30	100.0	2.3	100.0	2.1	100.0	1.6

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 22

Results of ω under MAR using FIML for Same Between-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
73	MAR	750	.050(B)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
74				15	100.0	-0.1	99.1	0.0	99.1	-3.9
75				30	100.0	-0.1	98.0	0.0	98.0	-3.7
76			.111(B)	0	100.0	-0.1	98.9	0.0	98.9	-0.4
77				15	100.0	0.0	97.3	0.1	97.3	-0.5
78				30	99.9	0.0	96.2	0.2	96.2	-0.8
79			.296(B)	0	96.2	-0.6	82.6	0.6	82.6	0.3
80				15	92.5	-0.7	76.3	1.1	76.3	0.3
81				30	90.3	-0.6	72.9	1.3	72.9	0.3
82		1500(G100)	.050(B)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
83				15	100.0	-0.1	99.6	0.0	99.6	-1.7
84				30	100.0	-0.1	99.8	0.0	99.8	-1.4
85			.111(B)	0	100.0	-0.1	100.0	-0.1	100.0	0.0
86				15	100.0	-0.1	99.8	-0.1	99.8	0.1
87				30	100.0	-0.1	99.9	-0.1	99.9	0.1
88			.296(B)	0	99.7	-0.6	97.9	-0.1	97.9	0.1
89				15	99.1	-0.9	96.3	0.1	96.3	0.1
90				30	99.2	-1.0	94.3	0.2	94.3	0.1
91		1500(G50)	.050(B)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
92				15	100.0	-0.1	99.9	0.0	99.9	-1.1
93				30	100.0	-0.1	99.8	0.0	99.8	-1.0
94			.111(B)	0	100.0	-0.1	99.9	-0.2	99.9	0.0
95				15	100.0	-0.1	100.0	-0.2	100.0	0.0
96				30	100.0	0.0	99.6	-0.1	99.6	0.1
97			.296(B)	0	99.5	-0.6	94.1	-0.1	94.1	0.1
98				15	99.0	-1.1	91.6	0.1	91.6	0.1
99				30	98.8	-0.8	90.2	0.4	90.2	0.3
100		3000	.050(B)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
101				15	100.0	-0.1	100.0	0.0	100.0	-0.3
102				30	100.0	0.0	100.0	0.0	100.0	-0.3
103			.111(B)	0	100.0	0.0	100.0	-0.1	100.0	0.0
104				15	100.0	0.0	100.0	-0.1	100.0	0.0
105				30	100.0	0.0	100.0	0.0	100.0	0.1
106			.296(B)	0	100.0	-0.4	99.8	-0.1	99.8	0.0
107				15	100.0	-0.8	99.6	0.0	99.6	0.0
108				30	100.0	-0.7	99.2	0.1	99.2	0.1

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 23

Results of ω under MAR using FIML for Same Within-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
109	MAR	750	.050(W)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
110				15	100.0	-0.1	99.1	0.0	99.1	-3.9
111				30	100.0	-0.1	100.0	1.5	100.0	-2.3
112			.111(W)	0	97.0	0.0	100.0	0.0	100.0	5.8
113				15	99.2	0.0	99.9	0.0	99.9	6.4
114				30	99.7	0.0	99.6	0.0	99.6	6.1
115			.296(W)	0	77.8	0.0	99.7	0.0	99.7	-1.8
116				15	86.8	0.0	100.0	0.0	100.0	-2.1
117				30	91.0	0.1	99.8	0.0	99.8	-2.0
118		1500(G100)	.050(W)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
119				15	100.0	-0.1	99.6	0.0	99.6	-1.7
120				30	100.0	-0.1	99.8	0.0	99.8	-1.4
121			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	1.1
122				15	100.0	-0.1	99.7	0.0	99.7	1.4
123				30	100.0	-0.1	99.9	0.0	99.9	1.5
124			.296(W)	0	99.2	-0.1	100.0	0.0	100.0	-0.6
125				15	99.7	-0.1	100.0	0.0	100.0	-0.7
126				30	99.7	0.0	100.0	0.0	100.0	-0.7
127		1500(G50)	.050(W)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
128				15	100.0	-0.1	99.9	0.0	99.9	-1.1
129				30	100.0	-0.1	99.8	0.0	99.8	-1.0
130			.111(W)	0	98.4	0.0	99.9	0.0	99.9	6.3
131				15	99.7	0.0	99.9	0.0	99.9	5.9
132				30	99.9	0.0	99.5	0.0	99.5	6.5
133			.296(W)	0	79.4	0.1	99.7	0.0	99.7	-0.9
134				15	88.5	0.1	100.0	0.0	100.0	-0.8
135				30	92.0	0.0	99.8	0.0	99.8	-0.8
136		3000	.050(W)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
137				15	100.0	-0.1	100.0	0.0	100.0	-0.3
138				30	100.0	0.0	100.0	0.0	100.0	-0.3
139			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	2.5
140				15	100.0	-0.1	99.9	0.0	99.9	2.5
141				30	100.0	0.0	100.0	0.0	100.0	2.3
142			.296(W)	0	99.4	0.0	99.8	0.0	99.8	-0.4
143				15	99.7	0.0	100.0	0.0	100.0	-0.3
144				30	99.8	0.0	100.0	0.0	100.0	-0.3

Note. #=condition ID regarding Table 6; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 24

Results of α under MNAR using LD for Same Between-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
145	MNAR	750	.050(B)	0	100.0	1.5	100.0	1.5	100.0	-1.5
146				15	100.0	-6.9	100.0	-6.6	100.0	-23.8
147				30	100.0	-11.5	100.0	-11.0	100.0	-29.2
148			.111(B)	0	100.0	2.7	100.0	2.8	100.0	-0.2
149				15	100.0	-11.1	100.0	-10.2	100.0	-17.0
150				30	100.0	-18.0	100.0	-17.0	100.0	-26.6
151			.296(B)	0	100.0	3.4	100.0	5.3	100.0	1.2
152				15	100.0	-15.3	100.0	-13.1	100.0	-5.7
153				30	100.0	-24.5	100.0	-21.8	100.0	-14.3
154		1500(G100)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
155				15	100.0	-6.9	100.0	-6.5	100.0	-26.8
156				30	100.0	-11.5	100.0	-11.0	100.0	-35.0
157			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.4
158				15	100.0	-11.1	100.0	-10.2	100.0	-14.4
159				30	100.0	-17.9	100.0	-16.9	100.0	-25.6
160			.296(B)	0	100.0	3.7	100.0	5.3	100.0	1.7
161				15	100.0	-15.0	100.0	-13.2	100.0	-4.3
162				30	100.0	-24.1	100.0	-22.1	100.0	-10.5
163		1500(G50)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
164				15	100.0	-6.9	100.0	-6.5	100.0	-27.7
165				30	100.0	-11.6	100.0	-11.0	100.0	-36.0
166			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.2
167				15	100.0	-11.1	100.0	-10.3	100.0	-13.2
168				30	100.0	-18.0	100.0	-16.9	100.0	-23.6
169			.296(B)	0	100.0	3.6	100.0	5.2	100.0	1.5
170				15	100.0	-15.2	100.0	-13.4	100.0	-4.1
171				30	100.0	-24.7	100.0	-22.5	100.0	-10.3
172		3000	.050(B)	0	100.0	1.6	100.0	1.5	100.0	1.2
173				15	100.0	-6.8	100.0	-6.4	100.0	-28.2
174				30	100.0	-11.4	100.0	-10.8	100.0	-39.9
175			.111(B)	0	100.0	2.8	100.0	2.8	100.0	1.7
176				15	100.0	-10.9	100.0	-10.1	100.0	-11.3
177				30	100.0	-17.6	100.0	-16.6	100.0	-20.9
178			.296(B)	0	100.0	4.0	100.0	5.6	100.0	1.8
179				15	100.0	-14.6	100.0	-13.0	100.0	-3.5
180				30	100.0	-23.4	100.0	-21.7	100.0	-8.3

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 25

Results of α under MNAR using LD for Same Within-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
181	MNAR	750	.050(W)	0	100.0	1.5	100.0	1.5	100.0	-1.5
182				15	100.0	-6.9	100.0	-6.6	100.0	-23.8
183				30	100.0	-11.5	100.0	-11.0	100.0	-29.2
184			.111(W)	0	100.0	2.0	100.0	1.5	100.0	5.8
185				15	100.0	-8.2	100.0	-5.6	100.0	-39.9
186				30	100.0	-13.4	100.0	-10.2	100.0	-38.3
187			.296(W)	0	100.0	2.3	100.0	2.0	100.0	-0.5
188				15	100.0	-8.4	100.0	-4.4	100.0	-38.4
189				30	100.0	-14.2	100.0	-8.3	100.0	-48.9
190		1500(G100)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
191				15	100.0	-6.9	100.0	-6.5	100.0	-26.8
192				30	100.0	-11.5	100.0	-11.0	100.0	-35.0
193			.111(W)	0	100.0	1.9	100.0	1.5	100.0	0.7
194				15	100.0	-8.2	100.0	-5.7	100.0	-58.7
195				30	100.0	-13.4	100.0	-10.1	100.0	-58.0
196			.296(W)	0	100.0	2.3	100.0	2.0	100.0	1.2
197				15	100.0	-8.4	100.0	-4.3	100.0	-39.0
198				30	100.0	-14.2	100.0	-8.2	100.0	-53.2
199		1500(G50)	.050(W)	0	100.0	1.5	100.0	1.5	100.0	0.1
200				15	100.0	-6.9	100.0	-6.5	100.0	-27.7
201				30	100.0	-11.6	100.0	-11.0	100.0	-36.0
202			.111(W)	0	100.0	2.0	100.0	1.5	100.0	3.1
203				15	100.0	-8.2	100.0	-5.6	100.0	-57.4
204				30	100.0	-13.4	100.0	-10.0	100.0	-61.0
205			.296(W)	0	100.0	2.3	100.0	2.0	100.0	0.4
206				15	100.0	-8.5	100.0	-4.3	100.0	-39.0
207				30	100.0	-14.2	100.0	-8.1	100.0	-52.8
208		3000	.050(W)	0	100.0	1.6	100.0	1.5	100.0	1.2
209				15	100.0	-6.8	100.0	-6.4	100.0	-28.2
210				30	100.0	-11.4	100.0	-10.8	100.0	-39.9
211			.111(W)	0	100.0	2.0	100.0	1.5	100.0	1.7
212				15	100.0	-8.1	100.0	-5.5	100.0	-69.1
213				30	100.0	-13.3	100.0	-9.6	100.0	-68.3
214			.296(W)	0	100.0	2.4	100.0	2.1	100.0	1.7
215				15	100.0	-8.3	100.0	-4.2	100.0	-38.2
216				30	100.0	-14.0	100.0	-8.0	100.0	-55.5

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 26

Results of ω under MNAR using LD for Same Between-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
145	MNAR	750	.050(B)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
146				15	100.0	-7.9	98.4	-7.4	98.4	-25.3
147				30	99.8	-11.9	93.6	-11.3	93.6	-25.1
148			.111(B)	0	100.0	-0.1	98.9	0.0	98.9	-0.4
149				15	98.1	-12.9	81.3	-11.6	81.3	-13.5
150				30	88.8	-18.4	63.7	-16.7	63.7	-19.6
151			.296(B)	0	96.2	-0.6	82.6	0.6	82.6	0.3
152				15	80.7	-17.0	54.5	-14.7	54.5	-4.4
153				30	68.2	-25.0	35.5	-20.6	35.5	-7.8
154		1500(G100)	.050(B)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
155				15	100.0	-7.9	99.3	-7.3	99.3	-29.4
156				30	100.0	-12.0	98.0	-11.3	98.0	-33.8
157			.111(B)	0	100.0	-0.1	100.0	-0.1	100.0	0.0
158				15	100.0	-13.0	97.2	-12.1	97.2	-13.1
159				30	98.6	-18.9	84.0	-17.6	84.0	-20.8
160			.296(B)	0	99.7	-0.6	97.9	-0.1	97.9	0.1
161				15	93.2	-17.7	80.5	-16.5	80.5	-4.3
162				30	80.1	-25.3	54.1	-23.2	54.1	-7.2
163		1500(G50)	.050(B)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
164				15	100.0	-7.9	98.9	-7.3	98.9	-29.2
165				30	100.0	-12.0	97.8	-11.3	97.8	-36.6
166			.111(B)	0	100.0	-0.1	99.9	-0.2	99.9	0.0
167				15	100.0	-13.0	98.0	-12.1	98.0	-12.4
168				30	98.0	-19.0	80.6	-17.6	80.6	-19.6
169			.296(B)	0	99.5	-0.6	94.1	-0.1	94.1	0.1
170				15	93.1	-17.7	74.0	-16.4	74.0	-3.9
171				30	77.9	-25.3	50.2	-22.6	50.2	-7.8
172		3000	.050(B)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
173				15	100.0	-7.8	99.2	-7.3	99.2	-28.8
174				30	100.0	-11.9	99.2	-11.2	99.2	-40.9
175			.111(B)	0	100.0	0.0	100.0	-0.1	100.0	0.0
176				15	100.0	-12.8	100.0	-12.1	100.0	-11.6
177				30	100.0	-18.8	97.9	-17.7	97.9	-18.6
178			.296(B)	0	100.0	-0.4	99.8	-0.1	99.8	0.0
179				15	99.0	-17.6	95.2	-17.0	95.2	-4.2
180				30	91.1	-25.0	78.0	-24.0	78.0	-7.1

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 27

Results of ω under MNAR using LD for Same Within-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
181	MNAR	750	.050$^{(W)}$	0	100.0	-0.1	99.0	0.0	99.0	-3.5
182				15	100.0	-7.9	98.4	-7.4	98.4	-25.3
183				30	99.8	-11.9	93.5	-11.3	93.5	-25.1
184			.111$^{(W)}$	0	97.0	0.0	100.0	0.0	100.0	5.8
185				15	95.9	-9.1	99.1	-6.7	99.1	-32.4
186				30	88.4	-13.3	94.1	-10.5	94.1	-26.3
187			.296$^{(W)}$	0	77.8	0.0	99.7	0.0	99.7	-1.8
188				15	74.3	-10.0	99.5	-6.4	99.5	-32.1
189				30	62.7	-14.8	97.0	-10.3	97.0	-38.6
190		1500$^{(G100)}$	.050$^{(W)}$	0	100.0	-0.1	99.7	0.0	99.7	-1.3
191				15	100.0	-7.9	99.3	-7.3	99.3	-29.4
192				30	100.0	-12.0	98.0	-11.3	98.0	-33.8
193			.111$^{(W)}$	0	100.0	-0.1	100.0	0.0	100.0	1.1
194				15	100.0	-9.3	99.7	-6.7	99.7	-43.1
195				30	99.3	-13.7	97.2	-10.5	97.2	-35.1
196			.296$^{(W)}$	0	99.2	-0.1	100.0	0.0	100.0	-0.6
197				15	98.4	-10.5	99.8	-6.4	99.8	-33.4
198				30	89.8	-15.7	98.6	-10.2	98.6	-41.9
199		1500$^{(G50)}$	.050$^{(W)}$	0	100.0	-0.1	99.9	0.0	99.9	-1.1
200				15	100.0	-7.9	98.9	-7.3	98.9	-29.2
201				30	100.0	-12.0	97.8	-11.3	97.8	-36.6
202			.111$^{(W)}$	0	98.4	0.0	99.9	0.0	99.9	6.3
203				15	98.6	-9.2	99.9	-6.6	99.9	-42.1
204				30	93.7	-13.7	99.0	-10.3	99.0	-38.5
205			.296$^{(W)}$	0	79.4	0.1	99.7	0.0	99.7	-0.9
206				15	77.5	-10.2	99.9	-6.4	99.9	-32.3
207				30	61.8	-15.2	99.5	-10.2	99.5	-42.0
208		3000	.050$^{(W)}$	0	100.0	-0.1	100.0	0.0	100.0	-0.3
209				15	100.0	-7.8	99.2	-7.3	99.2	-28.8
210				30	100.0	-11.9	99.2	-11.2	99.2	-40.9
211			.111$^{(W)}$	0	100.0	-0.1	100.0	0.0	100.0	2.5
212				15	100.0	-9.2	100.0	-6.5	100.0	-49.1
213				30	99.8	-13.6	99.8	-10.1	99.8	-46.5
214			.296$^{(W)}$	0	99.4	0.0	99.8	0.0	99.8	-0.4
215				15	99.1	-10.5	100.0	-6.4	100.0	-33.1
216				30	93.7	-15.6	99.8	-10.0	99.8	-43.8

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 28

Results of α under MNAR using FIML for Same Between-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
145	MNAR	750	.050(B)	0	100.0	1.5	100.0	1.5	100.0	-1.5
146				15	100.0	-2.1	100.0	-2.1	100.0	-4.9
147				30	100.0	-4.6	100.0	-4.6	100.0	-8.2
148			.111(B)	0	100.0	2.7	100.0	2.8	100.0	-0.2
149				15	100.0	-5.3	100.0	-5.5	100.0	-1.8
150				30	100.0	-10.1	100.0	-10.4	100.0	-4.6
151			.296(B)	0	100.0	3.4	100.0	5.3	100.0	1.2
152				15	100.0	-10.0	100.0	-11.9	100.0	0.1
153				30	100.0	-17.4	100.0	-20.8	100.0	-2.2
154		1500(G100)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
155				15	100.0	-2.1	100.0	-2.0	100.0	-2.8
156				30	100.0	-4.6	100.0	-4.6	100.0	-5.7
157			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.4
158				15	100.0	-5.3	100.0	-5.6	100.0	0.2
159				30	100.0	-10.0	100.0	-10.5	100.0	-2.2
160			.296(B)	0	100.0	3.7	100.0	5.3	100.0	1.7
161				15	100.0	-9.9	100.0	-12.0	100.0	0.7
162				30	100.0	-17.1	100.0	-21.0	100.0	-1.4
163		1500(G50)	.050(B)	0	100.0	1.5	100.0	1.5	100.0	0.1
164				15	100.0	-2.1	100.0	-2.0	100.0	-2.6
165				30	100.0	-4.6	100.0	-4.5	100.0	-5.4
166			.111(B)	0	100.0	2.7	100.0	2.7	100.0	1.2
167				15	100.0	-5.3	100.0	-5.5	100.0	0.1
168				30	100.0	-10.1	100.0	-10.5	100.0	-2.1
169			.296(B)	0	100.0	3.6	100.0	5.2	100.0	1.5
170				15	100.0	-10.0	100.0	-11.5	100.0	0.4
171				30	100.0	-17.5	100.0	-20.7	100.0	-1.7
172		3000	.050(B)	0	100.0	1.6	100.0	1.5	100.0	1.2
173				15	100.0	-2.0	100.0	-2.0	100.0	-1.2
174				30	100.0	-4.6	100.0	-4.5	100.0	-3.6
175			.111(B)	0	100.0	2.8	100.0	2.8	100.0	1.7
176				15	100.0	-5.1	100.0	-5.4	100.0	0.7
177				30	100.0	-9.9	100.0	-10.3	100.0	-1.0
178			.296(B)	0	100.0	4.0	100.0	5.6	100.0	1.8
179				15	100.0	-9.5	100.0	-11.1	100.0	0.8
180				30	100.0	-16.9	100.0	-20.1	100.0	-1.2

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 29

Results of α under MNAR using FIML for Same Within-Group Level Parameters Conditions

					α		α_w		α_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
181	MNAR	750	.050$^{(W)}$	0	100.0	1.5	100.0	1.5	100.0	-1.5
182				15	100.0	-2.1	100.0	-2.1	100.0	-4.9
183				30	100.0	-4.6	100.0	-4.6	100.0	-8.2
184			.111$^{(W)}$	0	100.0	2.0	100.0	1.5	100.0	5.8
185				15	100.0	-2.8	100.0	-1.6	100.0	2.2
186				30	100.0	-6.0	100.0	-4.1	100.0	-4.0
187			.296$^{(W)}$	0	100.0	2.3	100.0	2.0	100.0	-0.5
188				15	100.0	-2.5	100.0	-0.5	100.0	-3.0
189				30	100.0	-6.0	100.0	-2.7	100.0	-6.6
190		1500$^{(G100)}$	.050$^{(W)}$	0	100.0	1.5	100.0	1.5	100.0	0.1
191				15	100.0	-2.1	100.0	-2.0	100.0	-2.8
192				30	100.0	-4.6	100.0	-4.6	100.0	-5.7
193			.111$^{(W)}$	0	100.0	1.9	100.0	1.5	100.0	0.7
194				15	100.0	-2.9	100.0	-1.6	100.0	-6.1
195				30	100.0	-6.1	100.0	-4.0	100.0	-12.5
196			.296$^{(W)}$	0	100.0	2.3	100.0	2.0	100.0	1.2
197				15	100.0	-2.6	100.0	-0.5	100.0	-1.3
198				30	100.0	-6.0	100.0	-2.6	100.0	-4.9
199		1500$^{(G50)}$	.050$^{(W)}$	0	100.0	1.5	100.0	1.5	100.0	0.1
200				15	100.0	-2.1	100.0	-2.0	100.0	-2.6
201				30	100.0	-4.6	100.0	-4.5	100.0	-5.4
202			.111$^{(W)}$	0	100.0	2.0	100.0	1.5	100.0	3.1
203				15	100.0	-2.8	100.0	-1.4	100.0	-1.8
204				30	100.0	-6.0	100.0	-3.7	100.0	-7.4
205			.296$^{(W)}$	0	100.0	2.3	100.0	2.0	100.0	0.4
206				15	100.0	-2.5	100.0	-0.3	100.0	-2.0
207				30	100.0	-6.0	100.0	-2.2	100.0	-5.3
208		3000	.050$^{(W)}$	0	100.0	1.6	100.0	1.5	100.0	1.2
209				15	100.0	-2.0	100.0	-2.0	100.0	-1.2
210				30	100.0	-4.6	100.0	-4.5	100.0	-3.6
211			.111$^{(W)}$	0	100.0	2.0	100.0	1.5	100.0	1.7
212				15	100.0	-2.9	100.0	-1.4	100.0	-5.1
213				30	100.0	-6.0	100.0	-3.6	100.0	-11.5
214			.296$^{(W)}$	0	100.0	2.4	100.0	2.1	100.0	1.7
215				15	100.0	-2.5	100.0	-0.2	100.0	-0.6
216				30	100.0	-6.0	100.0	-2.2	100.0	-3.7

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); α=single-level coefficient alpha; α_w=within-group level coefficient alpha; α_B=between-group level coefficient alpha.

Table 30

Results of ω under MNAR using FIML for Same Between-Group Level Parameters Conditions

					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
145	MNAR	750	.050(B)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
146				15	100.0	-2.3	98.5	-2.2	98.5	-5.8
147				30	100.0	-3.6	98.8	-3.5	98.8	-7.7
148			.111(B)	0	100.0	-0.1	98.9	0.0	98.9	-0.4
149				15	99.9	-6.6	94.2	-6.9	94.2	-0.4
150				30	99.4	-10.0	90.1	-10.4	90.1	-1.0
151			.296(B)	0	96.2	-0.6	82.6	0.6	82.6	0.3
152				15	89.1	-12.2	69.3	-14.1	69.3	0.4
153				30	84.0	-18.0	57.4	-21.3	57.4	-0.2
154		1500(G100)	.050(B)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
155				15	100.0	-2.3	99.4	-2.2	99.4	-3.8
156				30	100.0	-3.6	99.5	-3.5	99.5	-5.7
157			.111(B)	0	100.0	-0.1	100.0	-0.1	100.0	0.0
158				15	100.0	-6.7	99.7	-7.0	99.7	-0.1
159				30	100.0	-10.1	99.1	-10.6	99.1	-0.8
160			.296(B)	0	99.7	-0.6	97.9	-0.1	97.9	0.1
161				15	98.7	-12.7	92.6	-15.5	92.6	0.0
162				30	95.4	-18.4	84.8	-22.9	84.8	-0.6
163		1500(G50)	.050(B)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
164				15	100.0	-2.3	99.4	-2.2	99.4	-3.1
165				30	100.0	-3.6	99.7	-3.5	99.7	-4.8
166			.111(B)	0	100.0	-0.1	99.9	-0.2	99.9	0.0
167				15	100.0	-6.7	99.8	-6.9	99.8	-0.1
168				30	100.0	-10.1	99.3	-10.6	99.3	-0.7
169			.296(B)	0	99.5	-0.6	94.1	-0.1	94.1	0.1
170				15	97.7	-12.6	85.3	-14.8	85.3	0.0
171				30	93.3	-18.6	76.4	-22.4	76.4	-0.6
172		3000	.050(B)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
173				15	100.0	-2.3	100.0	-2.2	100.0	-2.2
174				30	100.0	-3.6	99.8	-3.5	99.8	-3.8
175			.111(B)	0	100.0	0.0	100.0	-0.1	100.0	0.0
176				15	100.0	-6.6	100.0	-6.9	100.0	-0.1
177				30	100.0	-10.0	100.0	-10.5	100.0	-0.7
178			.296(B)	0	100.0	-0.4	99.8	-0.1	99.8	0.0
179				15	100.0	-12.5	99.4	-15.1	99.4	-0.2
180				30	99.9	-18.5	97.8	-22.7	97.8	-0.9

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Table 31

Results of ω under MNAR using FIML for Same Within-Group Level Parameters Conditions

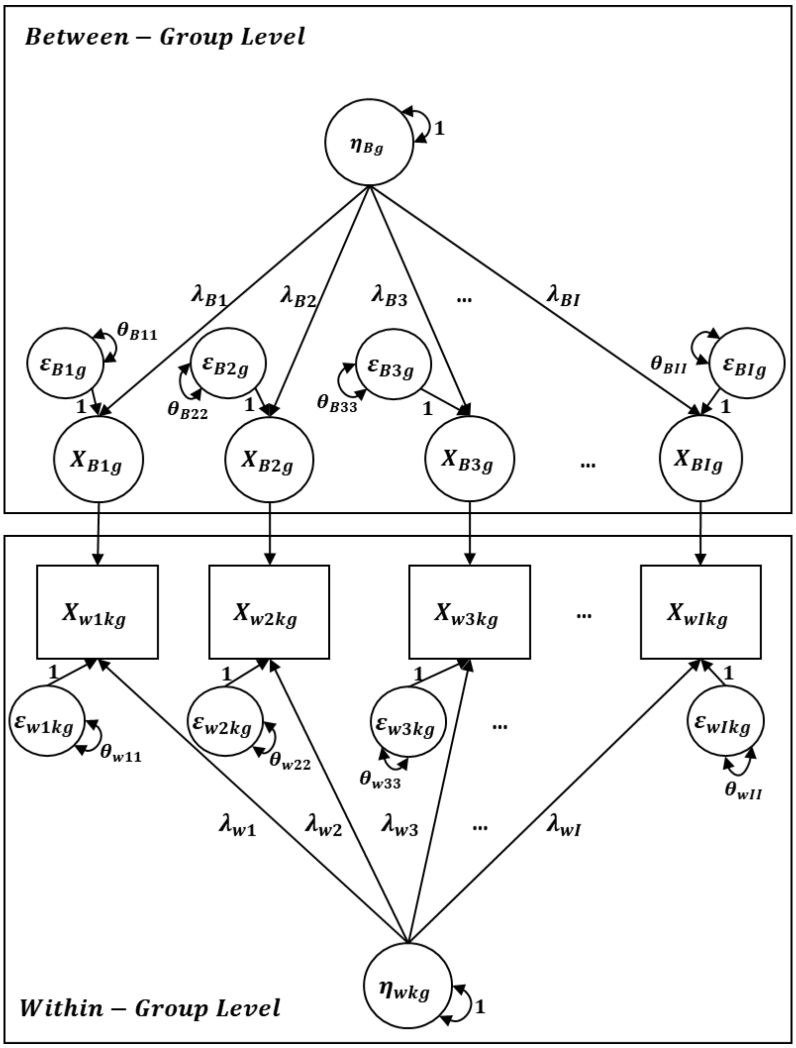
					ω		ω_W		ω_B	
#	MDM	SS	ICC	MDP	CR	PB	CR	PB	CR	PB
181	MNAR	750	.050(W)	0	100.0	-0.1	99.0	0.0	99.0	-3.5
182				15	100.0	-2.3	98.5	-2.2	98.5	-5.8
183				30	100.0	-3.6	98.7	-3.5	98.7	-7.6
184			.111(W)	0	97.0	0.0	100.0	0.0	100.0	5.8
185				15	99.6	-3.2	99.8	-2.0	99.8	3.3
186				30	99.5	-4.9	99.8	-3.4	99.8	-0.2
187			.296(W)	0	77.8	0.0	99.7	0.0	99.7	-1.8
188				15	90.9	-4.1	99.8	-2.3	99.8	-3.2
189				30	92.2	-6.5	99.5	-4.1	99.5	-5.1
190		1500(G100)	.050(W)	0	100.0	-0.1	99.7	0.0	99.7	-1.3
191				15	100.0	-2.3	99.4	-2.2	99.4	-3.8
192				30	100.0	-3.6	99.5	-3.5	99.5	-5.7
193			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	1.1
194				15	100.0	-3.3	99.9	-2.0	99.9	-3.5
195				30	100.0	-5.0	99.6	-3.4	99.6	-7.6
196			.296(W)	0	99.2	-0.1	100.0	0.0	100.0	-0.6
197				15	99.9	-4.3	99.9	-2.3	99.9	-2.0
198				30	100.0	-6.8	99.9	-4.1	99.9	-4.1
199		1500(G50)	.050(W)	0	100.0	-0.1	99.9	0.0	99.9	-1.1
200				15	100.0	-2.3	99.4	-2.2	99.4	-3.1
201				30	100.0	-3.6	99.7	-3.5	99.7	-4.8
202			.111(W)	0	98.4	0.0	99.9	0.0	99.9	6.3
203				15	99.8	-3.2	100.0	-1.9	100.0	2.3
204				30	100.0	-5.0	99.7	-3.1	99.7	-3.0
205			.296(W)	0	79.4	0.1	99.7	0.0	99.7	-0.9
206				15	94.1	-4.2	100.0	-2.2	100.0	-2.1
207				30	94.7	-6.6	100.0	-3.8	100.0	-3.9
208		3000	.050(W)	0	100.0	-0.1	100.0	0.0	100.0	-0.3
209				15	100.0	-2.3	100.0	-2.2	100.0	-2.2
210				30	100.0	-3.6	99.8	-3.5	99.8	-3.8
211			.111(W)	0	100.0	-0.1	100.0	0.0	100.0	2.5
212				15	100.0	-3.3	100.0	-1.9	100.0	-2.4
213				30	100.0	-5.0	99.6	-3.1	99.6	-7.3
214			.296(W)	0	99.4	0.0	99.8	0.0	99.8	-0.4
215				15	100.0	-4.3	100.0	-2.2	100.0	-1.5
216				30	100.0	-6.8	100.0	-3.7	100.0	-3.3

Note. #=condition ID regarding Table 7; MDM=missing data mechanisms; SS=sample sizes; ICC=intra-class correlations; MDP=missing data proportions (%); CR=convergence rate (%); PB=percentage bias (%); ω=single-level coefficient omega; ω_W=within-group level coefficient omega; ω_B=between-group level coefficient omega.

Figures

Figure 1

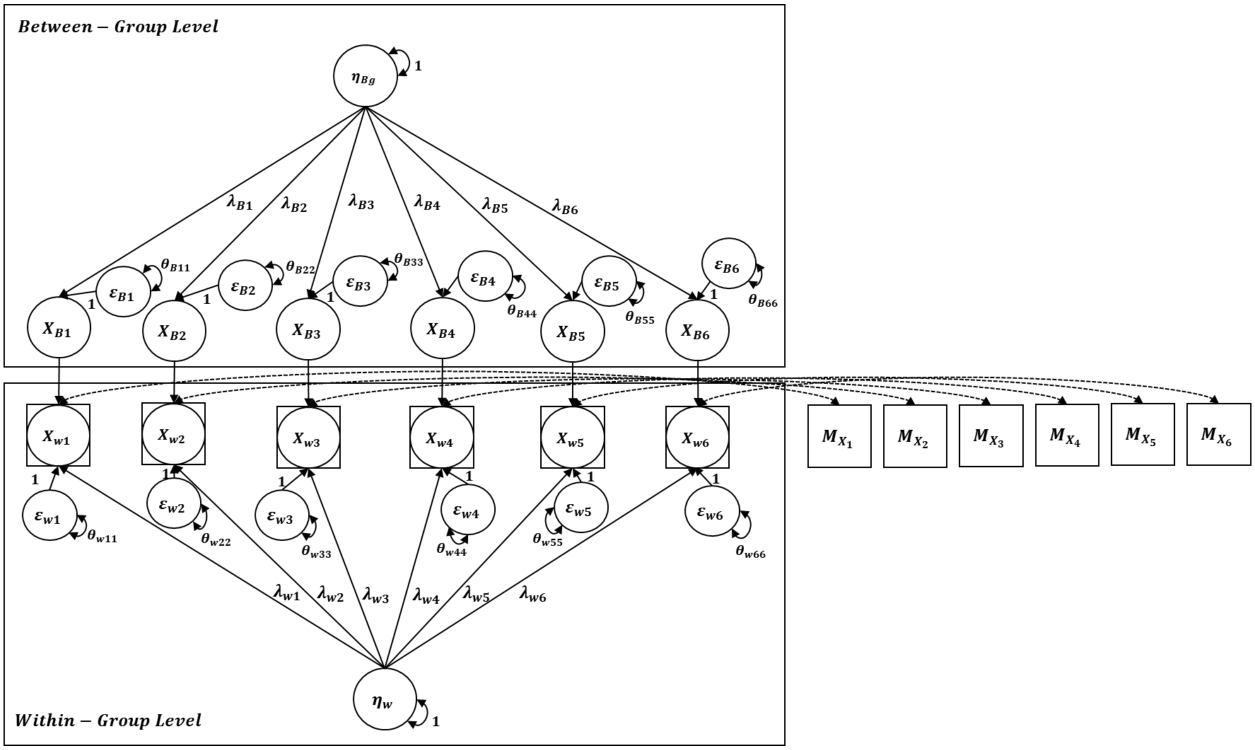
MCFA Model with Single Factor and I Indicators (Revised from Figure 1 in Muthén (1994))



Note. Lower plate=Within-group level model where X_{wikg} is the within-group level component of X_{ikg} decomposed by the factor loading λ_{wi} of the ith indicator, the factor score η_{wkg} of the kth examinee in the gth group, and the error ε_{wikg} of the ith indicator for the kth examinee in the gth group. Upper plate=Between-group level model where X_{Big} is the between-group level component of the observed score X_{ikg} decomposed by the factor loading λ_{Big} of the ith indicator, the factor score η_{Bg} of the kth examinee in the gth group, and the error ε_{Big} of the ith test item for the kth examinee in the gth group.

Figure 2

Model for Data Generation



Note. Lower plate=Within-group level model with a single factor represented with the factor score η_w in the circle linked to the six indicators X_i ($i = 1, \ldots, 6$) in the squares with the factor loadings λ_{wi} as the solid lines and their corresponding errors ε_{wi} in the circles with the error variances θ_{wii}. M_{X_i} in the squares indicated whether the corresponding indicator X_i was observed with their probabilities represented as the dashed lines. Upper plate=Between-group level model with a single factor represented as the circle of the factor score η_B linked with the six latent variables X_{Bi} ($i = 1, \ldots, 6$) in the circles with factor loadings λ_{Bi} as the solid lines and their corresponding errors ε_{Bi} in the circles with the error variances θ_{Bii}.

Figure 3

Missing Data Generation Process under MCAR for X_4 to X_6 with 30% Missingness

X1	X2	X3	X4	X5	X6
1.006	0.606	0.542	0.377	-0.208	-0.389
-0.612	1.236	0.205	-1.021	1.065	-0.424
-0.463	-0.935	-0.637	-0.940	-1.560	0.327
-0.453	-0.573	1.267	0.032	-1.283	0.072
1.445	1.324	1.149	0.486	0.381	1.328
2.785	0.720	1.013	2.008	0.779	1.965
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.715	0.987	0.578	0.018	1.074	0.968
0.789	-0.851	-0.285	0.831	-1.076	0.806
0.838	1.963	1.447	1.360	0.250	0.544
0.112	-0.788	0.588	0.179	-1.491	0.315
-0.742	0.673	0.377	0.721	-0.474	0.459
-0.713	-0.185	-0.568	-0.595	-0.270	-1.342
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.180	1.042	-0.518	-0.615	-0.496	-0.118
1.320	-0.123	0.512	1.124	-0.025	0.410
1.068	0.319	0.297	-0.607	-0.568	-0.340
-0.407	0.520	0.747	-1.045	-0.102	0.193
-0.090	-0.753	1.001	-0.945	-0.729	-0.521
-0.522	1.842	-0.690	-0.141	0.986	-0.412

<Raw Data>

→

X1	X2	X3	b4	X4	b5	X5	b6	X6
1.006	0.606	0.542	0	0.377	0	-0.208	0	-0.389
-0.612	1.236	0.205	1	NA	0	1.065	1	NA
-0.463	-0.935	-0.637	0	-0.940	1	NA	0	0.327
-0.453	-0.573	1.267	0	0.032	1	NA	0	0.072
1.445	1.324	1.149	1	NA	0	0.381	1	NA
2.785	0.720	1.013	0	2.008	0	0.779	0	1.965
-0.168	1.230	-0.169	0	-0.119	1	NA	0	0.901
-0.715	0.987	0.578	0	0.018	1	NA	1	NA
0.789	-0.851	-0.285	1	NA	0	-1.076	0	0.806
0.838	1.963	1.447	0	1.360	0	0.250	1	NA
0.112	-0.788	0.588	0	0.179	1	NA	0	0.315
-0.742	0.673	0.377	0	0.721	0	-0.474	0	0.459
-0.713	-0.185	-0.568	0	-0.595	0	-0.270	0	-1.342
-0.393	-0.317	0.625	1	NA	0	-0.268	1	NA
-0.180	1.042	-0.518	1	NA	0	-0.496	0	-0.118
1.320	-0.123	0.512	0	1.124	0	-0.025	1	NA
1.068	0.319	0.297	0	-0.607	0	-0.568	0	-0.340
-0.407	0.520	0.747	0	-1.045	0	-0.102	0	0.193
-0.090	-0.753	1.001	0	-0.945	1	NA	0	-0.521
-0.522	1.842	-0.690	1	NA	0	0.986	0	-0.412

<Data with Missing Values>

Figure 4

Missing Data Generation Process under MAR for X_4 with 30% Missingness

X1	X2	X3	X4	X5	X6
1.006	0.606	0.542	0.377	-0.208	-0.389
-0.612	1.236	0.205	-1.021	1.065	-0.424
-0.463	-0.935	-0.637	-0.940	-1.560	0.327
-0.453	-0.573	1.267	0.032	-1.283	0.072
1.445	1.324	1.149	0.486	0.381	1.328
2.785	0.720	1.013	2.008	0.779	1.965
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.715	0.987	0.578	0.018	1.074	0.968
0.789	-0.851	-0.285	0.831	-1.076	0.806
0.838	1.963	1.447	1.360	0.250	0.544
0.112	-0.788	0.588	0.179	-1.491	0.315
-0.742	0.673	0.377	0.721	-0.474	0.459
-0.713	-0.185	-0.568	-0.595	-0.270	-1.342
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.180	1.042	-0.518	-0.615	-0.496	-0.118
1.320	-0.123	0.512	1.124	-0.025	0.410
1.068	0.319	0.297	-0.607	-0.568	-0.340
-0.407	0.520	0.747	-1.045	-0.102	0.193
-0.090	-0.753	1.001	-0.945	-0.729	-0.521
-0.522	1.842	-0.690	-0.141	0.986	-0.412

<Raw Data>

→

X1	X2	X3	X4	X5	X6
-0.742	0.673	0.377	0.721	-0.474	0.459
-0.715	0.987	0.578	0.018	1.074	0.968
-0.713	-0.185	-0.568	-0.595	-0.270	-1.342
-0.612	1.236	0.205	-1.021	1.065	-0.424
-0.522	1.842	-0.690	-0.141	0.986	-0.412
-0.463	-0.935	-0.637	-0.940	-1.560	0.327
-0.453	-0.573	1.267	0.032	-1.283	0.072
-0.407	0.520	0.747	-1.045	-0.102	0.193
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.180	1.042	-0.518	-0.615	-0.496	-0.118
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.090	-0.753	1.001	-0.945	-0.729	-0.521
0.112	-0.788	0.588	0.179	-1.491	0.315
0.789	-0.851	-0.285	0.831	-1.076	0.806
0.838	1.963	1.447	1.360	0.250	0.544
1.006	0.606	0.542	0.377	-0.208	-0.389
1.068	0.319	0.297	-0.607	-0.568	-0.340
1.320	-0.123	0.512	1.124	-0.025	0.410
1.445	1.324	1.149	0.486	0.381	1.328
2.785	0.720	1.013	2.008	0.779	1.965

<Sorting X1>

→

X1	X2	X3	X4	X5	X6
-0.742	0.673	0.377	NA	-0.474	0.459
-0.715	0.987	0.578	NA	1.074	0.968
-0.713	-0.185	-0.568	NA	-0.270	-1.342
-0.612	1.236	0.205	NA	1.065	-0.424
-0.522	1.842	-0.690	NA	0.986	-0.412
-0.463	-0.935	-0.637	NA	-1.560	0.327
-0.453	-0.573	1.267	0.032	-1.283	0.072
-0.407	0.520	0.747	-1.045	-0.102	0.193
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.180	1.042	-0.518	-0.615	-0.496	-0.118
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.090	-0.753	1.001	-0.945	-0.729	-0.521
0.112	-0.788	0.588	0.179	-1.491	0.315
0.789	-0.851	-0.285	0.831	-1.076	0.806
0.838	1.963	1.447	1.360	0.250	0.544
1.006	0.606	0.542	0.377	-0.208	-0.389
1.068	0.319	0.297	-0.607	-0.568	-0.340
1.320	-0.123	0.512	1.124	-0.025	0.410
1.445	1.324	1.149	0.486	0.381	1.328
2.785	0.720	1.013	2.008	0.779	1.965

<30% Missing Values of X4>

Figure 5

Missing Data Generation Process under MNAR for X_4 with 30% Missingness

X1	X2	X3	X4	X5	X6
1.006	0.606	0.542	0.377	-0.208	-0.389
-0.612	1.236	0.205	-1.021	1.065	-0.424
-0.463	-0.935	-0.637	-0.940	-1.560	0.327
-0.453	-0.573	1.267	0.032	-1.283	0.072
1.445	1.324	1.149	0.486	0.381	1.328
2.785	0.720	1.013	2.008	0.779	1.965
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.715	0.987	0.578	0.018	1.074	0.968
0.789	-0.851	-0.285	0.831	-1.076	0.806
0.838	1.963	1.447	1.360	0.250	0.544
0.112	-0.788	0.588	0.179	-1.491	0.315
-0.742	0.673	0.377	0.721	-0.474	0.459
-0.713	-0.185	-0.568	-0.595	-0.270	-1.342
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.180	1.042	-0.518	-0.615	-0.496	-0.118
1.320	-0.123	0.512	1.124	-0.025	0.410
1.068	0.319	0.297	-0.607	-0.568	-0.340
-0.407	0.520	0.747	-1.045	-0.102	0.193
-0.090	-0.753	1.001	-0.945	-0.729	-0.521
-0.522	1.842	-0.690	-0.141	0.986	-0.412

<Raw Data>

→

X1	X2	X3	X4	X5	X6
-0.407	0.520	0.747	-1.045	-0.102	0.193
-0.612	1.236	0.205	-1.021	1.065	-0.424
-0.090	-0.753	1.001	-0.945	-0.729	-0.521
-0.463	-0.935	-0.637	-0.940	-1.560	0.327
-0.180	1.042	-0.518	-0.615	-0.496	-0.118
1.068	0.319	0.297	-0.607	-0.568	-0.340
-0.713	-0.185	-0.568	-0.595	-0.270	-1.342
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.522	1.842	-0.690	-0.141	0.986	-0.412
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.715	0.987	0.578	0.018	1.074	0.968
-0.453	-0.573	1.267	0.032	-1.283	0.072
0.112	-0.788	0.588	0.179	-1.491	0.315
1.006	0.606	0.542	0.377	-0.208	-0.389
1.445	1.324	1.149	0.486	0.381	1.328
-0.742	0.673	0.377	0.721	-0.474	0.459
0.789	-0.851	-0.285	0.831	-1.076	0.806
1.320	-0.123	0.512	1.124	-0.025	0.410
0.838	1.963	1.447	1.360	0.250	0.544
2.785	0.720	1.013	2.008	0.779	1.965

<Sorting X4>

→

X1	X2	X3	X4	X5	X6
-0.407	0.520	0.747	NA	-0.102	0.193
-0.612	1.236	0.205	NA	1.065	-0.424
-0.090	-0.753	1.001	NA	-0.729	-0.521
-0.463	-0.935	-0.637	NA	-1.560	0.327
-0.180	1.042	-0.518	NA	-0.496	-0.118
1.068	0.319	0.297	NA	-0.568	-0.340
-0.713	-0.185	-0.568	-0.595	-0.270	-1.342
-0.393	-0.317	0.625	-0.222	-0.268	-0.699
-0.522	1.842	-0.690	-0.141	0.986	-0.412
-0.168	1.230	-0.169	-0.119	-0.946	0.901
-0.715	0.987	0.578	0.018	1.074	0.968
-0.453	-0.573	1.267	0.032	-1.283	0.072
0.112	-0.788	0.588	0.179	-1.491	0.315
1.006	0.606	0.542	0.377	-0.208	-0.389
1.445	1.324	1.149	0.486	0.381	1.328
-0.742	0.673	0.377	0.721	-0.474	0.459
0.789	-0.851	-0.285	0.831	-1.076	0.806
1.320	-0.123	0.512	1.124	-0.025	0.410
0.838	1.963	1.447	1.360	0.250	0.544
2.785	0.720	1.013	2.008	0.779	1.965

<30% Missing Values of X4>

Figure 6

Convergence Rates of Coefficient Omega under MCAR using LD Method

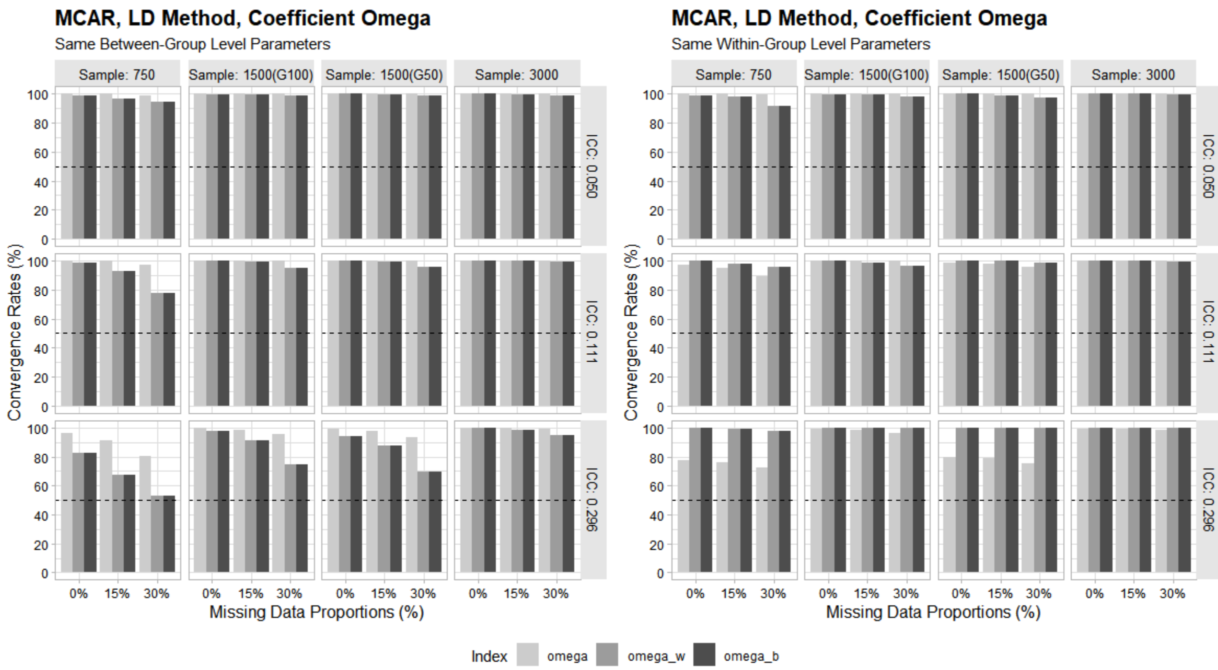


Figure 7

Percentage bias of Coefficient Alpha under MCAR using LD Method

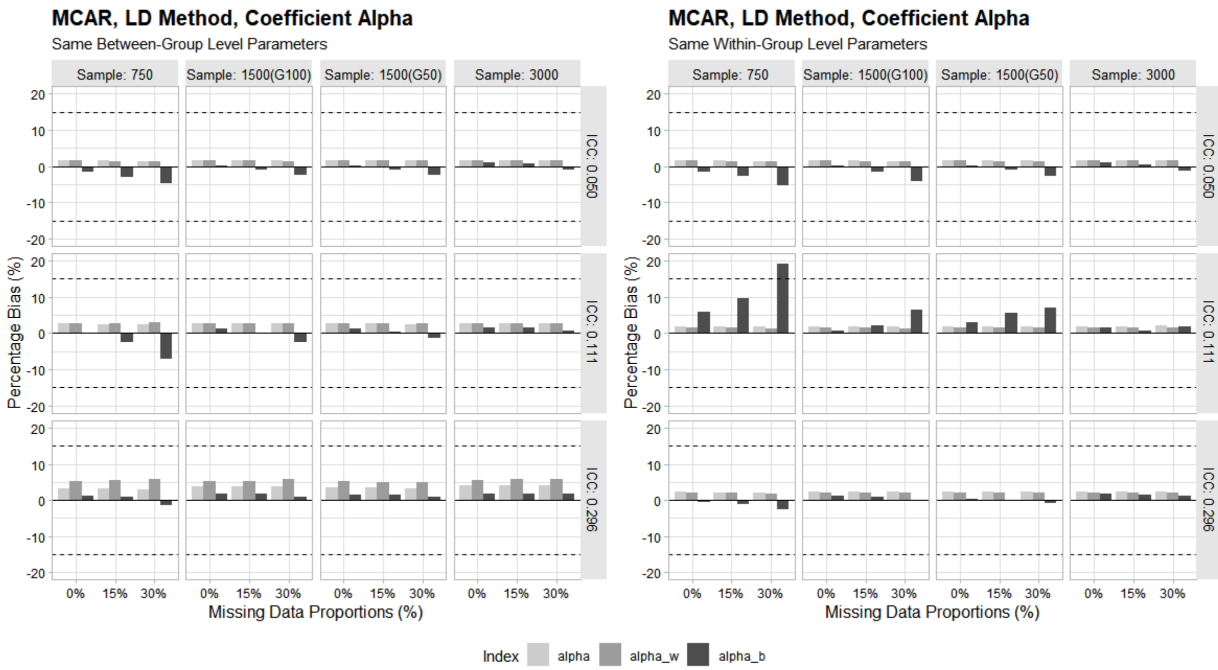


Figure 8

Percentage bias of Coefficient Omega under MCAR using LD Method

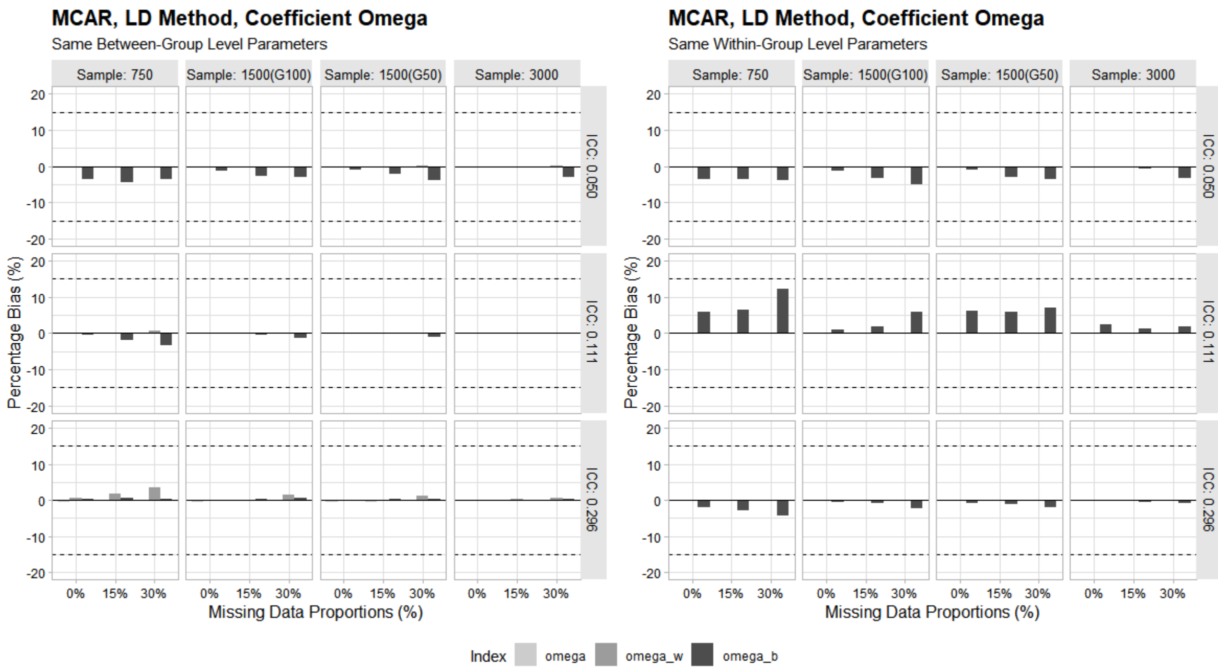


Figure 9

Convergence Rates of Coefficient Omega under MCAR using FIML Method

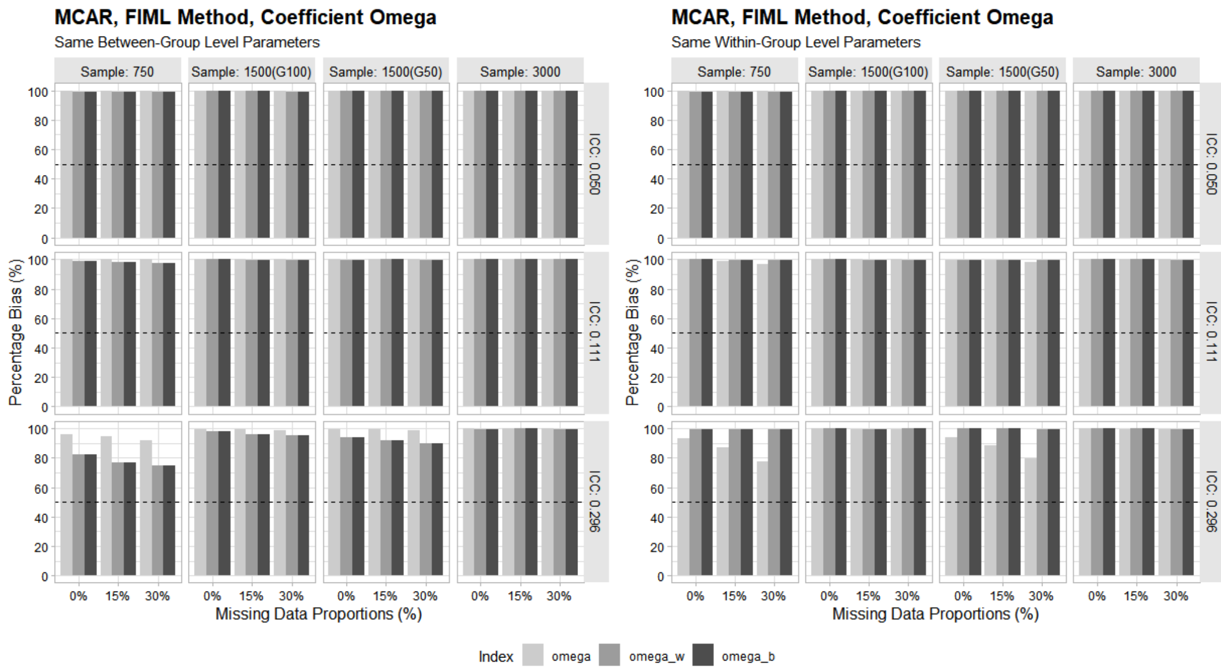


Figure 10

Percentage bias of Coefficient Alpha under MCAR using FIML Method

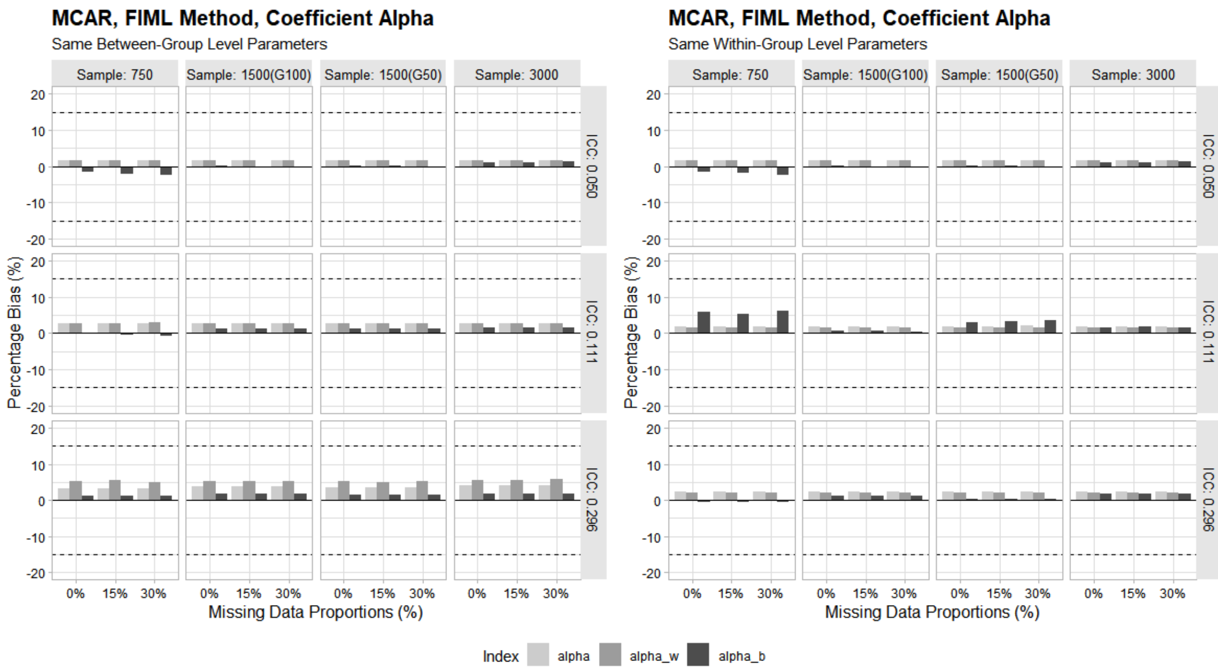


Figure 11

Percentage bias of Coefficient Omega under MCAR using FIML Method

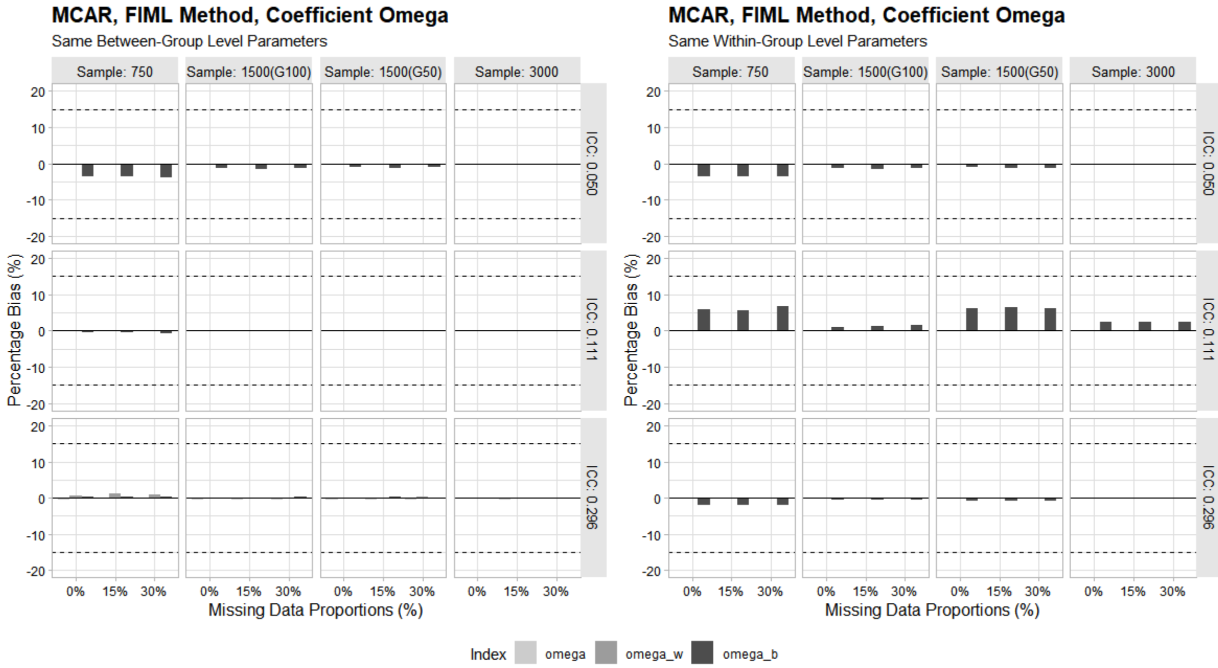


Figure 12

Convergence Rates of Coefficient Omega under MAR using LD Method

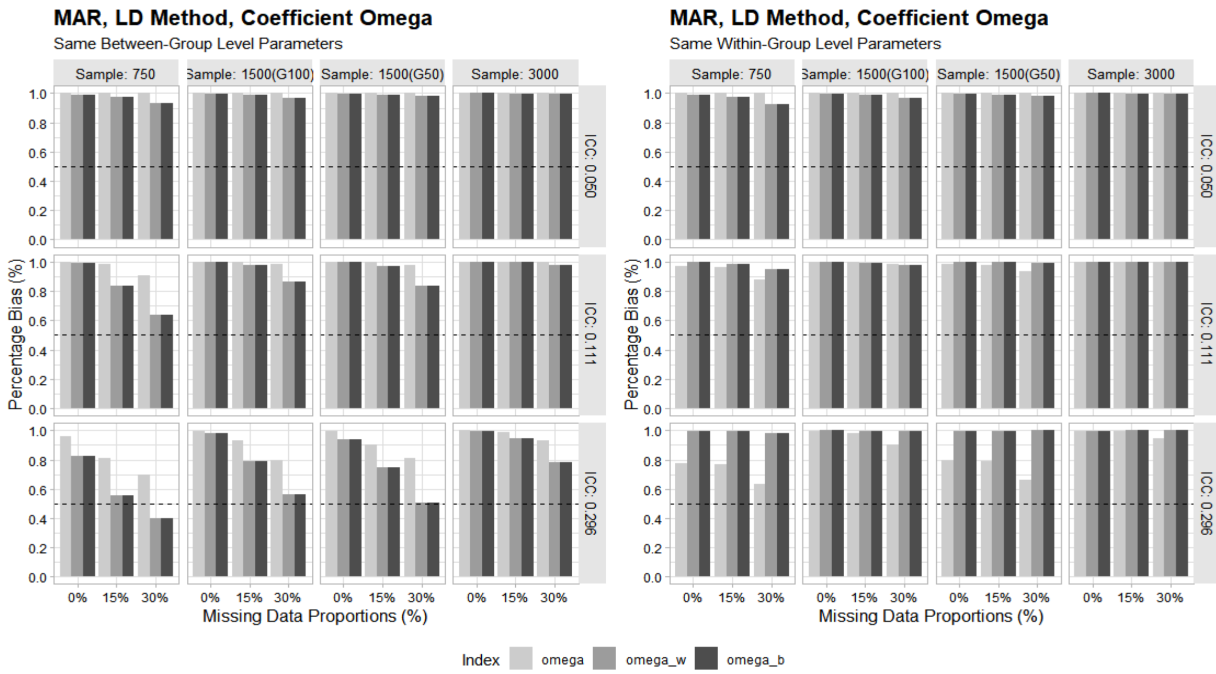


Figure 13

Percentage bias of Coefficient Alpha under MAR using LD Method

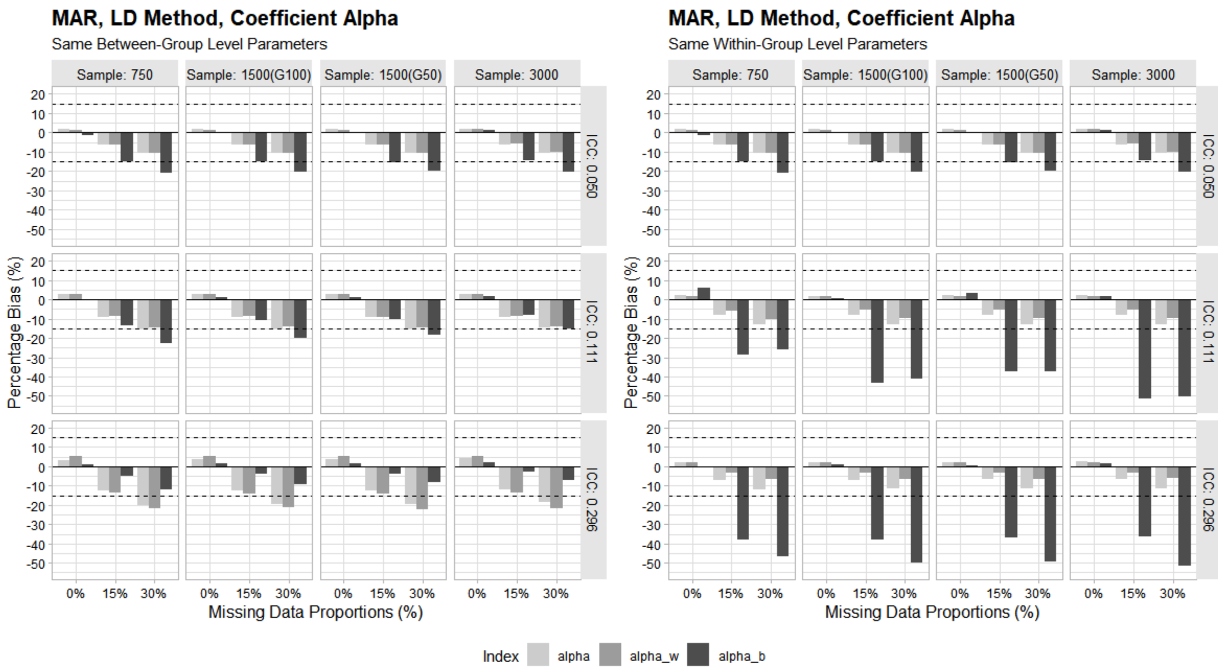


Figure 14

Percentage bias of Coefficient Omega under MAR using LD Method

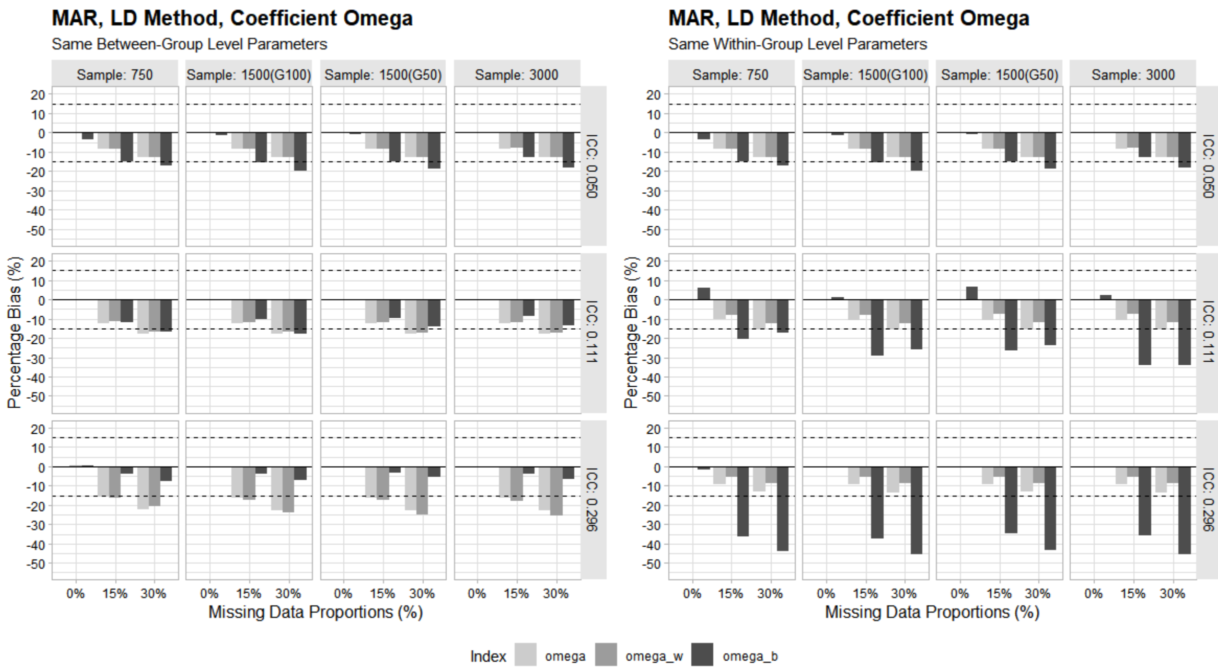


Figure 15

Convergence Rates of Coefficient Omega under MAR using FIML Method

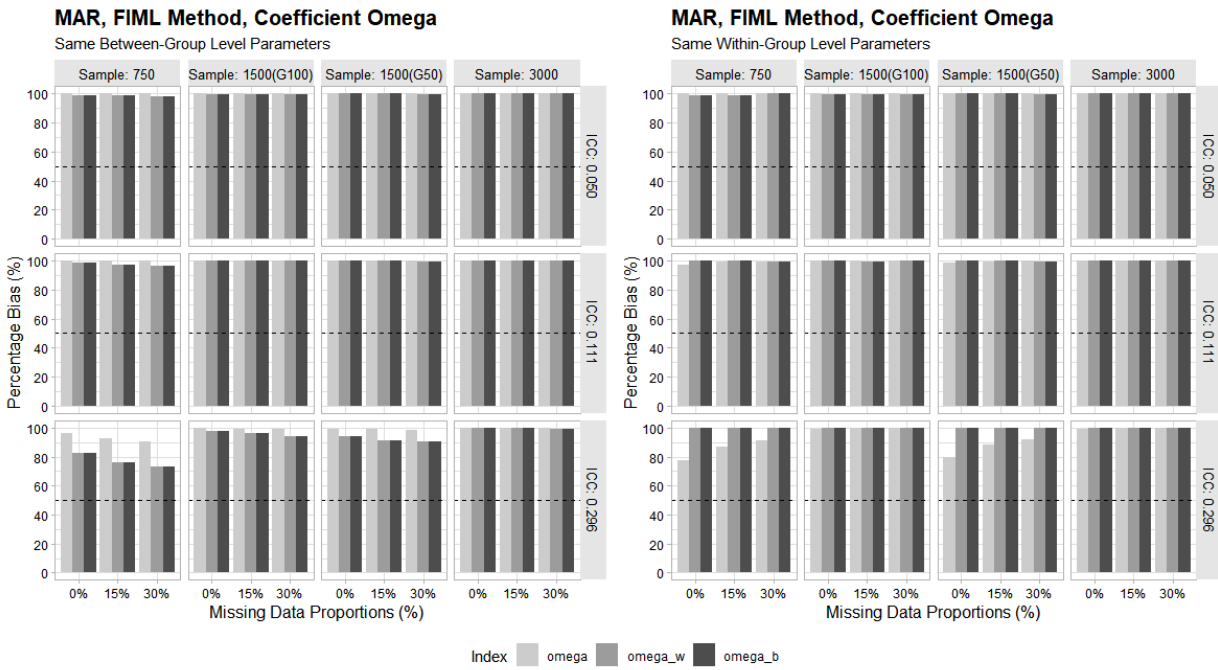


Figure 16

Percentage bias of Coefficient Alpha under MAR using FIML Method

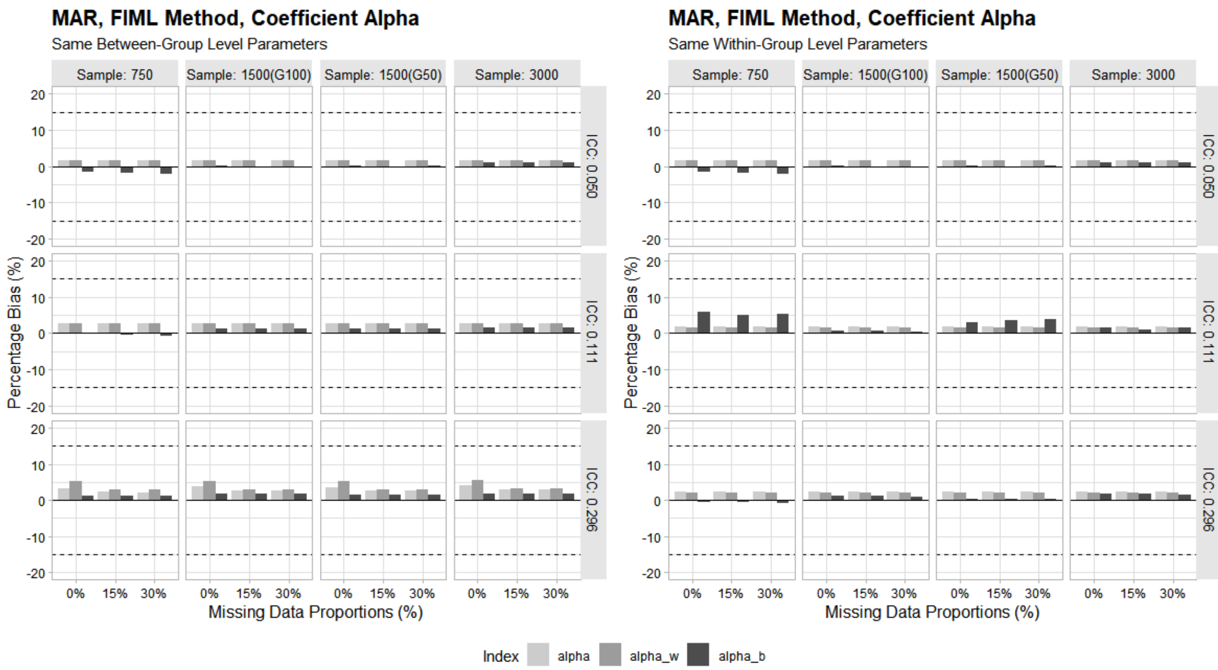


Figure 17

Percentage bias of Coefficient Omega under MAR using FIML Method

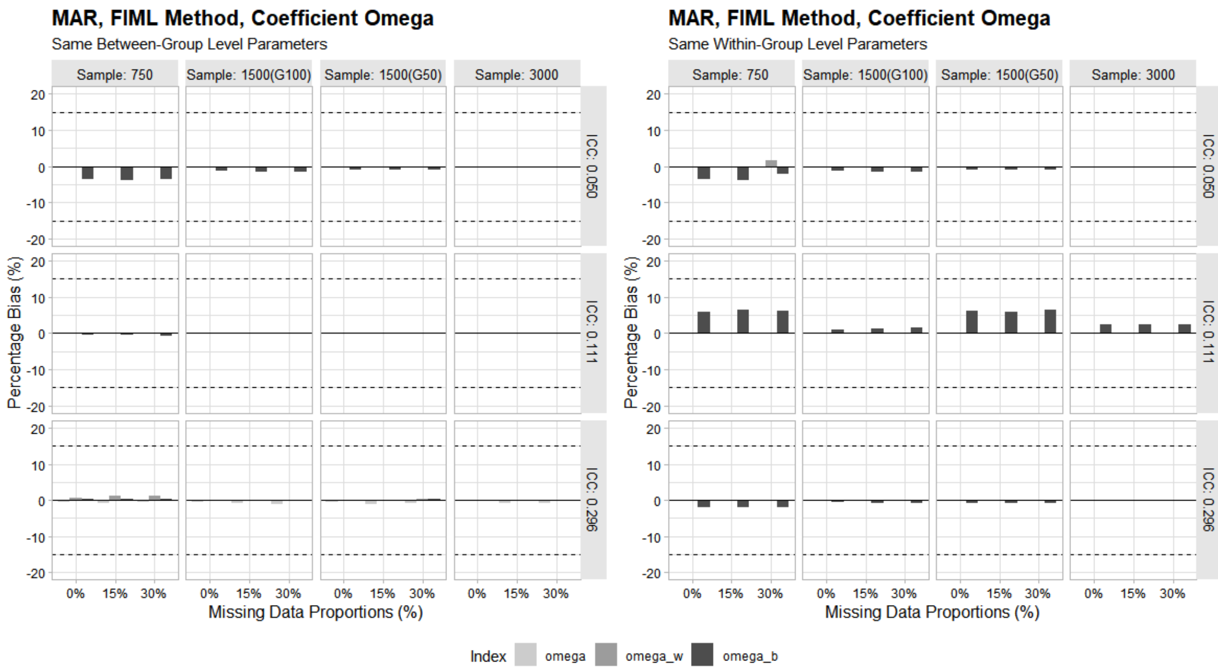


Figure 18

Convergence Rates of Coefficient Omega under MNAR using LD Method

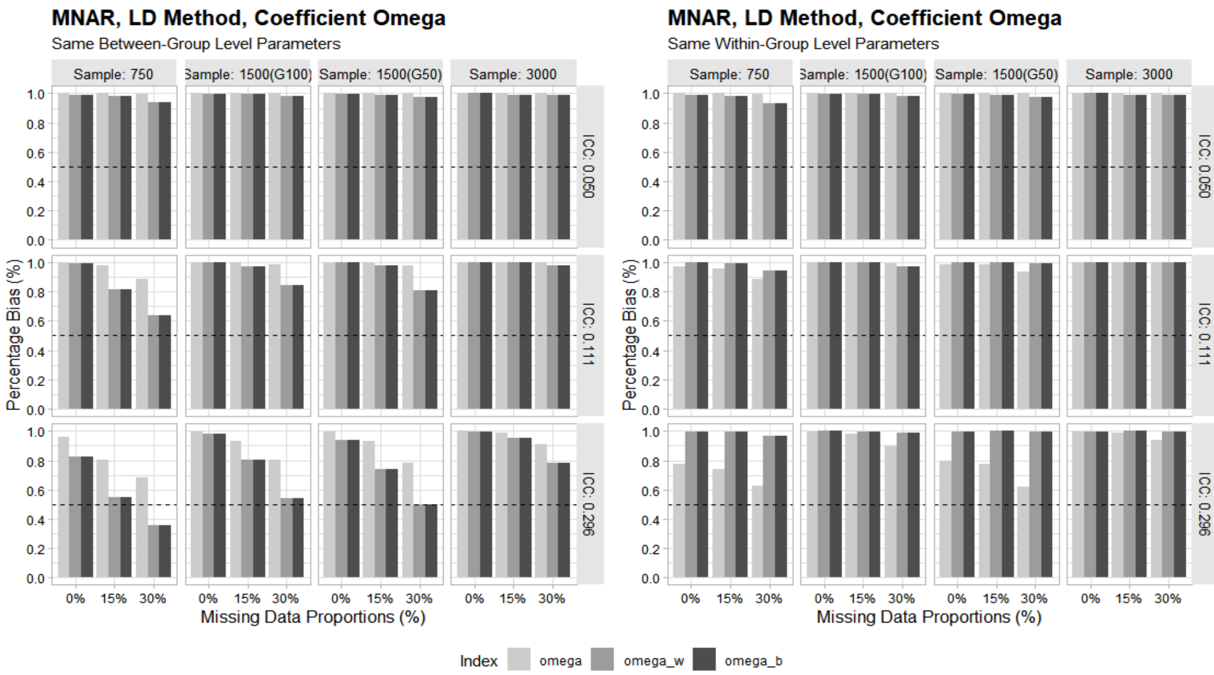


Figure 19

Percentage bias of Coefficient Alpha under MNAR using LD Method

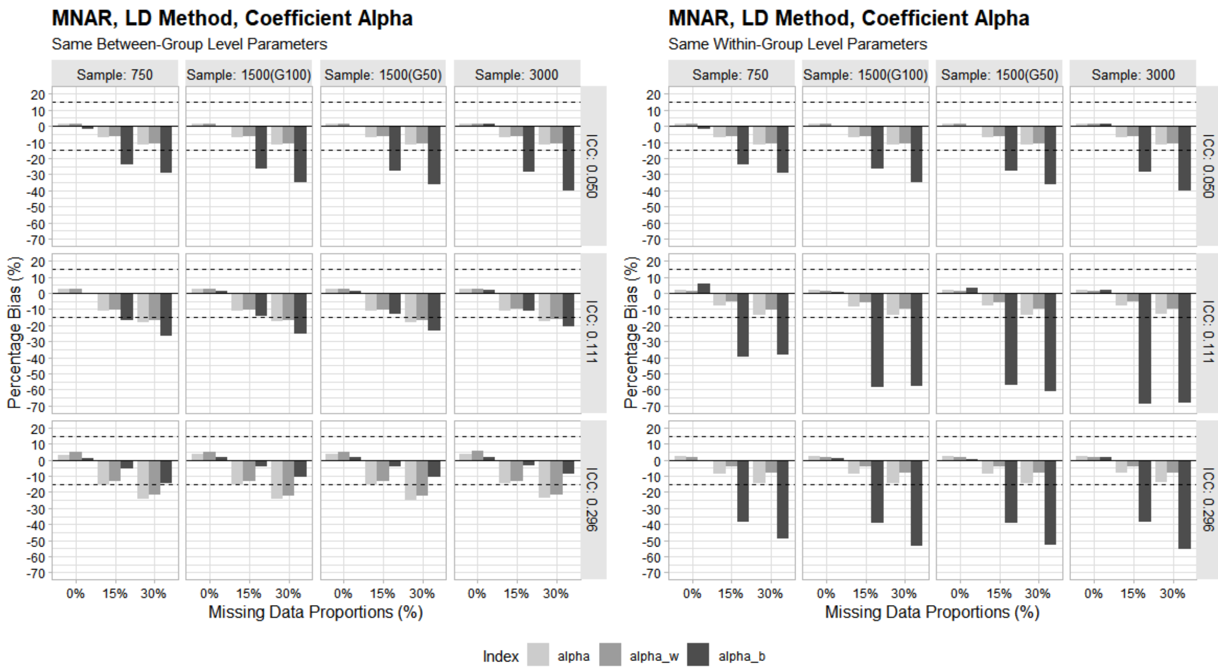


Figure 20

Percentage bias of Coefficient Omega under MNAR using LD Method

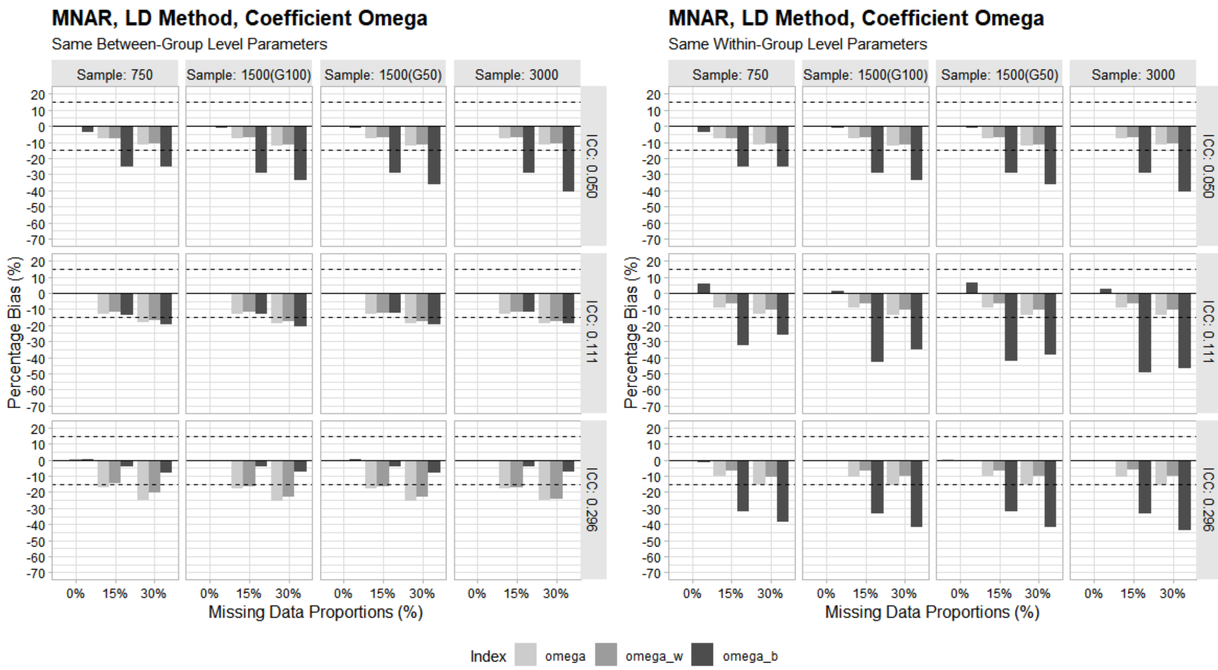


Figure 21

Convergence Rates of Coefficient Omega under MNAR using FIML Method

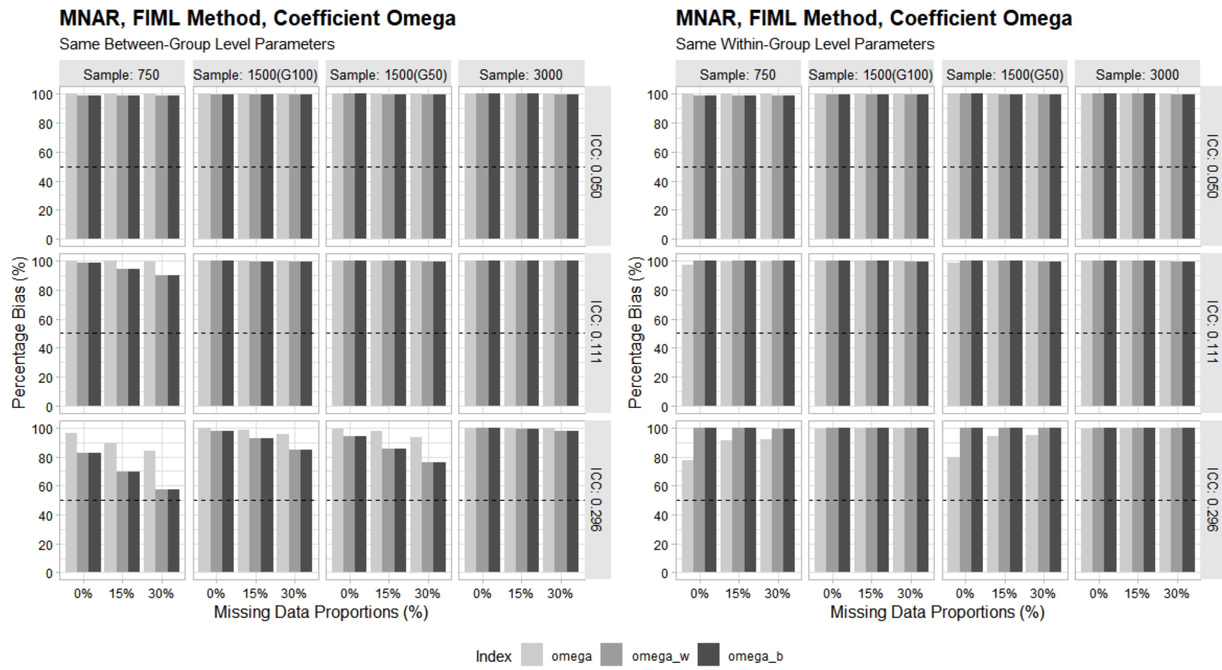


Figure 22

Percentage bias of Coefficient Alpha under MNAR using FIML Method

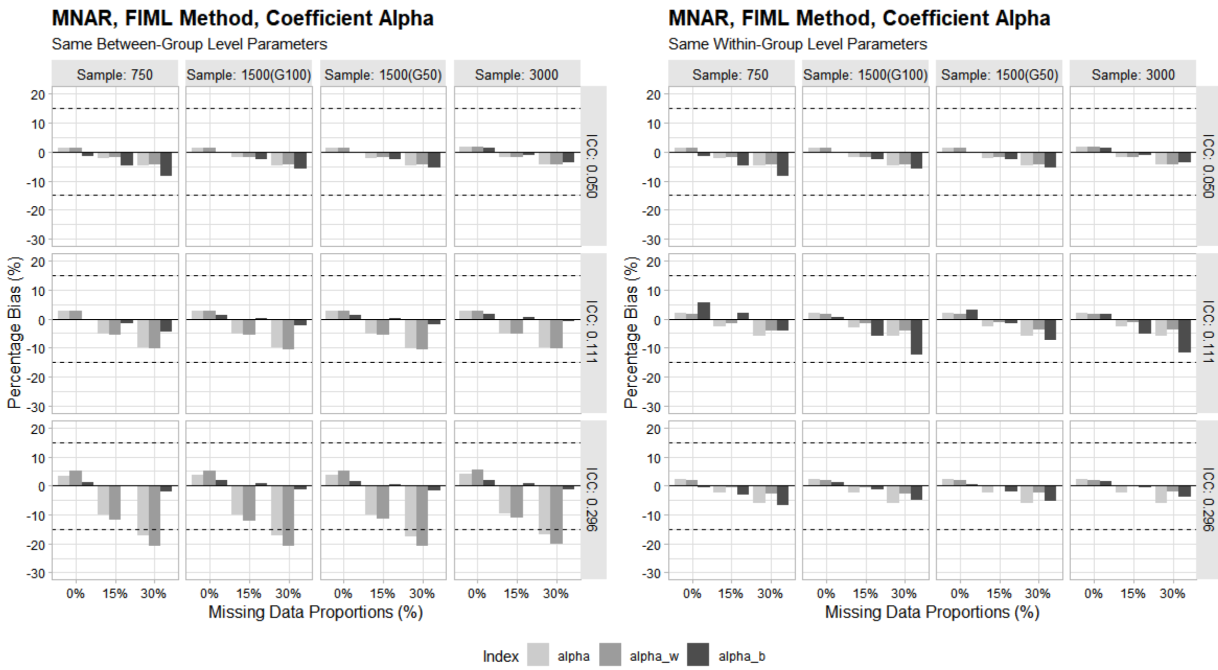
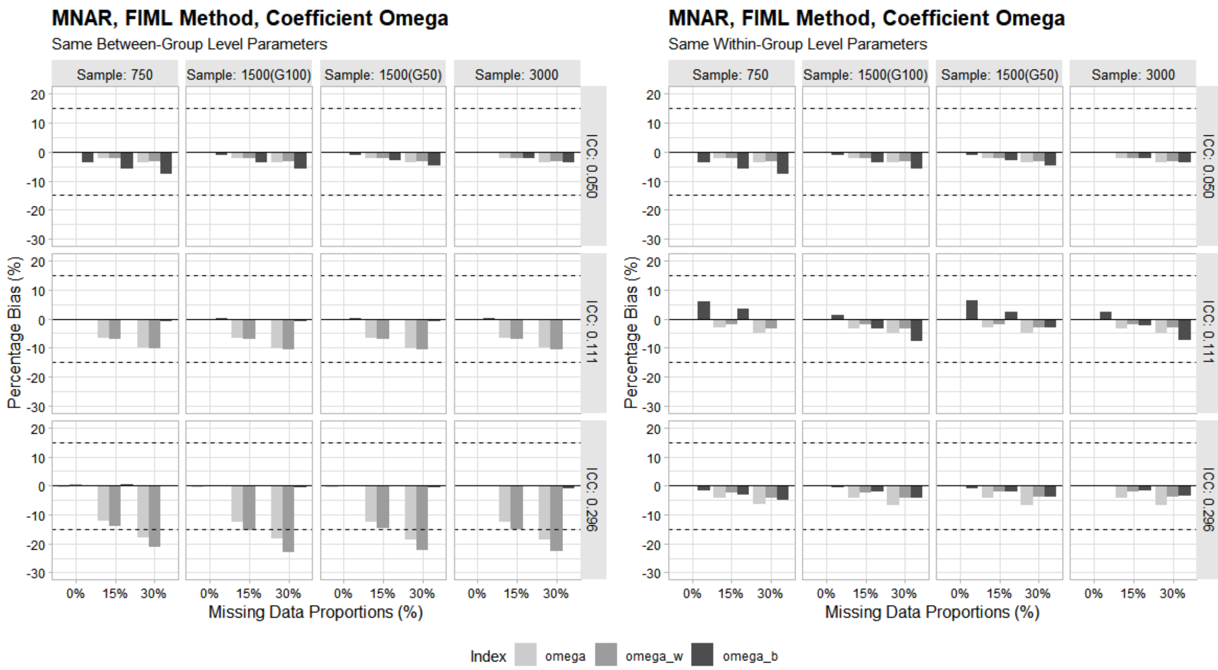


Figure 23

Percentage bias of Coefficient Omega under MNAR using FIML Method



Appendices

Appendix 1. *Mplus* code for single-level alpha α using LD method

```
TITLE: single-level alpha;
DATA: FILE = trial_1.dat;
   LISTWISE=ON;
VARIABLE: NAMES = X1-X6 CLASS;
   USEVARIABLES = X1-X6;
   MISSING = ALL(999);
MODEL:
   X1 WITH X2 (c1);
   X1 WITH X3 (c2);
   X1 WITH X4 (c3);
   X1 WITH X5 (c4);
   X1 WITH X6 (c5);
   X2 WITH X3 (c6);
   X2 WITH X4 (c7);
   X2 WITH X5 (c8);
   X2 WITH X6 (c9);
   X3 WITH X4 (c10);
   X3 WITH X5 (c11);
   X3 WITH X6 (c12);
   X4 WITH X5 (c13);
   X4 WITH X6 (c14);
   X5 WITH X6 (c15);
   X1 (v1);
   X2 (v2);
   X3 (v3);
   X4 (v4);
   X5 (v5);
   X6 (v6);
OUTPUT: SAMPSTAT;
MODEL CONSTRAINT: NEW(COMP_V ALPHA);
COMP_V =
v1+v2+v3+v4+v5+v6+2*(c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c13+c14+c15);
ALPHA =
(((c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c13+c14+c15)/15)*36)/COMP_V;
SAVEDATA:
RESULTS = res_trial_1_single_level_alpha.dat;
SAMPLE = sam_trial_1_single_level_alpha.dat;
```

Appendix 2. *Mplus* code for single-level omega ω using LD method

```
TITLE: single-level omega;
DATA: FILE = trial_1.dat;
    LISTWISE=ON;
VARIABLE: NAMES = X1-X6 CLASS;
    USEVARIABLES = X1-X6;
    MISSING = ALL(999);
MODEL:
    f BY X1* (l1);
    f BY X2 (l2);
   f BY X3 (l3);
   f BY X4 (l4);
   f BY X5 (l5);
   f BY X6 (l6);
   f@1;
   X1 (r1);
   X2 (r2);
   X3 (r3);
   X4 (r4);
   X5 (r5);
   X6 (r6);
OUTPUT: SAMPSTAT;
MODEL CONSTRAINT: NEW(NUM DENOM OMEGA);
   NUM = (l1+l2+l3+l4+l5+l6)**2;
   DENOM = ((l1+l2+l3+l4+l5+l6)**2)+(r1+r2+r3+r4+r5+r6);
   OMEGA = NUM/DENOM;
SAVEDATA:
   RESULTS = res_trial_1_single_level_omega.dat;
   SAMPLE = sam_trial_1_single_level_omega.dat;
```

Appendix 3. *Mplus* code for level-specific alpha α_W and α_B using LD method

```
TITLE: level-specific alpha;
DATA: FILE = trial_1.dat;
   LISTWISE=ON;
VARIABLE: NAMES = X1-X6 CLASS;
   USEVARIABLES = X1-X6;
   CLUSTER = CLASS;
   MISSING = ALL(999);
ANALYSIS: TYPE=TWOLEVEL;
MODEL:
%WITHIN%
   X1 WITH X2 (wc1);  X1 WITH X3 (wc2); X1 WITH X4 (wc3);
   X1 WITH X5 (wc4);  X1 WITH X6 (wc5);
   X2 WITH X3 (wc6);  X2 WITH X4 (wc7); X2 WITH X5 (wc8);
   X2 WITH X6 (wc9);
   X3 WITH X4 (wc10);    X3 WITH X5 (wc11);    X3 WITH X6 (wc12);
   X4 WITH X5 (wc13);    X4 WITH X6 (wc14);    X5 WITH X6 (wc15);
   X1 (wv1);    X2 (wv2);   X3 (wv3);    X4 (wv4);    X5 (wv5);   X6 (wv6);
%BETWEEN%
   X1 WITH X2 (bc1);  X1 WITH X3 (bc2);  X1 WITH X4 (bc3);
   X1 WITH X5 (bc4);  X1 WITH X6 (bc5);
   X2 WITH X3 (bc6);  X2 WITH X4 (bc7);  X2 WITH X5 (bc8);
   X2 WITH X6 (bc9);
   X3 WITH X4 (bc10);   X3 WITH X5 (bc11);    X3 WITH X6 (bc12);
   X4 WITH X5 (bc13);   X4 WITH X6 (bc14);    X5 WITH X6 (bc15);
   X1 (bv1);   X2 (bv2);   X3 (bv3);    X4 (bv4);    X5 (bv5);    X6 (bv6);
OUTPUT: SAMPSTAT;
MODEL CONSTRAINT: NEW(COMP_V_W ALPHAW COMP_V_B ALPHAB);
   COMP_V_W =
wv1+wv2+wv3+wv4+wv5+wv6+2*(wc1+wc2+wc3+wc4+wc5+wc6+wc7+wc8+wc9+wc10
  +wc11+wc12+wc13+wc13+wc14+wc15);
   ALPHAW = (((wc1+wc2+wc3+wc4+wc5+wc6+wc7+wc8+wc9+wc10
+wc11+wc12+wc13+wc13+wc14+wc15)/15)*36)/COMP_V_W;
   COMP_V_B =
bv1+bv2+bv3+bv4+bv5+bv6+2*(bc1+bc2+bc3+bc4+bc5+bc6+bc7+bc8+bc9+bc10
   +bc11+bc12+bc13+bc13+bc14+bc15);
   ALPHAB = (((bc1+bc2+bc3+bc4+bc5+bc6+bc7+bc8+bc9+bc10
+bc11+bc12+bc13+bc13+bc14+bc15)/15)*36)/COMP_V_B;
SAVEDATA: RESULTS = res_trial_1_two_level_alpha.dat;
   sample = sam_trial_1_two_level_alpha.dat;
```

Appendix 4. *Mplus* code for level-specific omega ω_W and ω_B using LD method

```
TITLE: level-specific omega;
DATA: FILE = trial_1.dat;
   LISTWISE=ON;
VARIABLE: NAMES = X1-X6 CLASS;
   USEVARIABLES = X1-X6;
   CLUSTER = CLASS;
   MISSING = ALL(999);
ANALYSIS: TYPE=TWOLEVEL;
MODEL:
 %WITHIN%
   fw BY X1* (wl1);
   fw BY X2 (wl2);
   fw BY X3 (wl3);
   fw BY X4 (wl4);
   fw BY X5 (wl5);
   fw BY X6 (wl6);
   fw@1;    X1 (wr1);    X2 (wr2);    X3 (wr3);    X4 (wr4);    X5 (wr5);    X6 (wr6);
   %BETWEEN%
   fb BY X1* (bl1);
   fb BY X2 (bl2);
   fb BY X3 (bl3);
   fb BY X4 (bl4);
   fb BY X5 (bl5);
   fb BY X6 (bl6);
   fb@1;    X1 (br1);    X2 (br2);    X3 (br3);    X4 (br4);    X5 (br5);    X6 (br6);
OUTPUT:    SAMPSTAT;
MODEL CONSTRAINT: NEW(NUMW DENOMW OMEGAW NUMB DENOMB
OMEGAB);
   NUMW = (wl1+wl2+wl3+wl4+wl5+wl6)**2;
   DENOMW = ((wl1+wl2+wl3+wl4+wl5+wl6)**2)+(wr1+wr2+wr3+wr4+wr5+wr6);
   OMEGAW = NUMW/DENOMW;
   NUMB = (bl1+bl2+bl3+bl4+bl5+bl6)**2;
   DENOMB = ((bl1+bl2+bl3+bl4+bl5+bl6)**2)+(br1+br2+br3+br4+br5+br6);
   OMEGAB = NUMB/DENOMB;
   wr1>0; br1>0;    wr2>0; br2>0;    wr3>0; br3>0;
   wr4>0; br4>0;    wr5>0; br5>0;    wr6>0; br6>0;
SAVEDATA:    RESULTS = res_trial_1_two_level_omega.dat;
   SAMPLE = sam_trial_1_two_level_omega.dat;
```

Appendix 5. *Mplus* code for single-level alpha α using FIML method

```
TITLE: single-level alpha;
DATA: FILE = trial_1.dat;
VARIABLE: NAMES = X1-X6 CLASS;
   USEVARIABLES = X1-X6;
   MISSING = ALL(999);
MODEL:
   X1 WITH X2 (c1);
   X1 WITH X3 (c2);
   X1 WITH X4 (c3);
   X1 WITH X5 (c4);
   X1 WITH X6 (c5);
   X2 WITH X3 (c6);
   X2 WITH X4 (c7);
   X2 WITH X5 (c8);
   X2 WITH X6 (c9);
   X3 WITH X4 (c10);
   X3 WITH X5 (c11);
   X3 WITH X6 (c12);
   X4 WITH X5 (c13);
   X4 WITH X6 (c14);
   X5 WITH X6 (c15);
   X1 (v1);
   X2 (v2);
   X3 (v3);
   X4 (v4);
   X5 (v5);
   X6 (v6);
OUTPUT: SAMPSTAT;
MODEL CONSTRAINT: NEW(COMP_V ALPHA);
COMP_V =
v1+v2+v3+v4+v5+v6+2*(c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c13+c14+c15);
ALPHA =
(((c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c13+c14+c15)/15)*36)/COMP_V;
SAVEDATA:
RESULTS = res_trial_1_single_level_alpha.dat;
SAMPLE = sam_trial_1_single_level_alpha.dat;
```

Appendix 6. *Mplus* code for single-level omega ω using FIML method

```
TITLE: single-level omega;
DATA: FILE = trial_1.dat;
VARIABLE: NAMES = X1-X6 CLASS;
    USEVARIABLES = X1-X6;
    MISSING = ALL(999);
MODEL:
    f BY X1* (l1);
    f BY X2 (l2);
   f BY X3 (l3);
   f BY X4 (l4);
   f BY X5 (l5);
   f BY X6 (l6);
   f@1;
   X1 (r1);
   X2 (r2);
   X3 (r3);
   X4 (r4);
   X5 (r5);
   X6 (r6);
OUTPUT: SAMPSTAT;
MODEL CONSTRAINT: NEW(NUM DENOM OMEGA);
   NUM = (l1+l2+l3+l4+l5+l6)**2;
   DENOM = ((l1+l2+l3+l4+l5+l6)**2)+(r1+r2+r3+r4+r5+r6);
   OMEGA = NUM/DENOM;
SAVEDATA:
   RESULTS = res_trial_1_single_level_omega.dat;
   SAMPLE = sam_trial_1_single_level_omega.dat;
```

Appendix 7. *Mplus* code for level-specific alpha α_W and α_B using FIML method

```
TITLE: level-specific alpha;
DATA: FILE = trial_1.dat;
VARIABLE: NAMES = X1-X6 CLASS;
   USEVARIABLES = X1-X6;
   MISSING = ALL(999);
   CLUSTER = CLASS;
ANALYSIS: TYPE=TWOLEVEL;
MODEL:
%WITHIN%
   X1 WITH X2 (wc1);  X1 WITH X3 (wc2); X1 WITH X4 (wc3);
   X1 WITH X5 (wc4);  X1 WITH X6 (wc5);
   X2 WITH X3 (wc6);  X2 WITH X4 (wc7); X2 WITH X5 (wc8);
   X2 WITH X6 (wc9);
   X3 WITH X4 (wc10);    X3 WITH X5 (wc11);    X3 WITH X6 (wc12);
   X4 WITH X5 (wc13);    X4 WITH X6 (wc14);    X5 WITH X6 (wc15);
   X1 (wv1);    X2 (wv2);   X3 (wv3);    X4 (wv4);    X5 (wv5);   X6 (wv6);
%BETWEEN%
   X1 WITH X2 (bc1);  X1 WITH X3 (bc2);  X1 WITH X4 (bc3);
   X1 WITH X5 (bc4);  X1 WITH X6 (bc5);
   X2 WITH X3 (bc6);  X2 WITH X4 (bc7);  X2 WITH X5 (bc8);
   X2 WITH X6 (bc9);
   X3 WITH X4 (bc10);   X3 WITH X5 (bc11);    X3 WITH X6 (bc12);
   X4 WITH X5 (bc13);   X4 WITH X6 (bc14);    X5 WITH X6 (bc15);
   X1 (bv1);   X2 (bv2);   X3 (bv3);    X4 (bv4);    X5 (bv5);    X6 (bv6);
OUTPUT: SAMPSTAT;
MODEL CONSTRAINT: NEW(COMP_V_W ALPHAW COMP_V_B ALPHAB);
   COMP_V_W =
wv1+wv2+wv3+wv4+wv5+wv6+2*(wc1+wc2+wc3+wc4+wc5+wc6+wc7+wc8+wc9+wc10
 +wc11+wc12+wc13+wc13+wc14+wc15);
   ALPHAW = (((wc1+wc2+wc3+wc4+wc5+wc6+wc7+wc8+wc9+wc10
+wc11+wc12+wc13+wc13+wc14+wc15)/15)*36)/COMP_V_W;
   COMP_V_B =
bv1+bv2+bv3+bv4+bv5+bv6+2*(bc1+bc2+bc3+bc4+bc5+bc6+bc7+bc8+bc9+bc10
  +bc11+bc12+bc13+bc13+bc14+bc15);
   ALPHAB = (((bc1+bc2+bc3+bc4+bc5+bc6+bc7+bc8+bc9+bc10
+bc11+bc12+bc13+bc13+bc14+bc15)/15)*36)/COMP_V_B;
SAVEDATA: RESULTS = res_trial_1_two_level_alpha.dat;
   sample = sam_trial_1_two_level_alpha.dat;
```

Appendix 8. *Mplus* code for level-specific omega ω_W and ω_B using FIML method

```
TITLE: level-specific omega;
DATA: FILE = trial_1.dat;
VARIABLE: NAMES = X1-X6 CLASS;
   USEVARIABLES = X1-X6;
   MISSING = ALL(999);
   CLUSTER = CLASS;
ANALYSIS: TYPE=TWOLEVEL;
MODEL:
 %WITHIN%
   fw BY X1* (wl1);
   fw BY X2 (wl2);
   fw BY X3 (wl3);
   fw BY X4 (wl4);
   fw BY X5 (wl5);
   fw BY X6 (wl6);
   fw@1;    X1 (wr1);    X2 (wr2);    X3 (wr3);    X4 (wr4);    X5 (wr5);    X6 (wr6);
   %BETWEEN%
   fb BY X1* (bl1);
   fb BY X2 (bl2);
   fb BY X3 (bl3);
   fb BY X4 (bl4);
   fb BY X5 (bl5);
   fb BY X6 (bl6);
   fb@1;    X1 (br1);    X2 (br2);    X3 (br3);    X4 (br4);    X5 (br5);    X6 (br6);
OUTPUT:    SAMPSTAT;
MODEL CONSTRAINT: NEW(NUMW DENOMW OMEGAW NUMB DENOMB
OMEGAB);
   NUMW = (wl1+wl2+wl3+wl4+wl5+wl6)**2;
   DENOMW = ((wl1+wl2+wl3+wl4+wl5+wl6)**2)+(wr1+wr2+wr3+wr4+wr5+wr6);
   OMEGAW = NUMW/DENOMW;
   NUMB = (bl1+bl2+bl3+bl4+bl5+bl6)**2;
   DENOMB = ((bl1+bl2+bl3+bl4+bl5+bl6)**2)+(br1+br2+br3+br4+br5+br6);
   OMEGAB = NUMB/DENOMB;
   wr1>0; br1>0;    wr2>0; br2>0;    wr3>0; br3>0;
   wr4>0; br4>0;    wr5>0; br5>0;    wr6>0; br6>0;
SAVEDATA:    RESULTS = res_trial_1_two_level_omega.dat;
   SAMPLE = sam_trial_1_two_level_omega.dat;
```