



Investigating the performance of level-specific fit indices in multilevel confirmatory factor analysis with dichotomous indicators: A Monte Carlo study

John J. H. Lin¹ · Hsien-Yuan Hsu²

Accepted: 24 October 2022 / Published online: 23 November 2022
© The Psychonomic Society, Inc. 2022

Abstract

We conducted a Monte Carlo study to examine the performance of level-specific χ^2 test statistics and fit regarding their capacity to determine model fit at specific levels in multilevel confirmatory factor analysis with dichotomous indicators. Five design factors—numbers of groups (NG), group size (GS), intra-class correlation (ICC), thresholds of dichotomous indicators (THR), and factor loadings (FL)—were considered in this study. According to our simulation results, we recommend that practitioners should be aware that the performance of between-level-specific (b-l-s) χ^2 and fit indices was mainly influenced by ICC and FL, followed by NG. At the same time, THR could slightly weigh in the performance of b-l-s fit indices in some conditions. Both b-l-s χ^2 and fit indices were more promising indicators to correctly indicate model fit when ICC or FL increased. A small to medium NG (50–100) might be sufficient for b-l-s χ^2 and fit indices only if both ICC and factor loadings were high, while in remaining conditions, an NG of 200 was needed. Moreover, practitioners could use within-level-specific (w-l-s) χ^2 and fit indices (except for $RMSEA_w$) along with traditional cut-off values to evaluate within-level models comprising dichotomous indicators. W-l-s χ^2 and fit indices were more promising to determine model fit when FL increased. THR had a slight impact and could weigh in the performance of χ^2_w , $RMSEA_w$, CFI_w , and $TLLI_w$. Unfortunately, $RMSEA_w$ was heavily affected by FL and THR and could determine model fit only when FL was high and THR was symmetric.

Keywords Multilevel structural equation modeling · Confirmatory factor analysis · Fit index · Model evaluation · Intraclass correlation

Introduction

Multilevel confirmatory factor analysis (MCFA), or multilevel measurement modeling, is an imperative statistical approach to validate latent constructs that underlie multiple item responses collected in multilevel settings (e.g., students nested within schools). MCFA has been widely used in psychological and educational research, and relevant reporting guidelines have been made available for substantive researchers (Kim et al., 2016). Unlike conventional

confirmatory factor analysis, the construct of interest in MCFA could be at multiple levels, resulting in challenges for substantive researchers in model specification, evaluation, and interpretation. Due to this complexity, extensive efforts have been devoted to providing instructions for appropriate specification and interpretation of MCFA (Stapleton, McNeish et al., 2016a; Stapleton, Yang et al., 2016b).

Regarding model evaluation, a body of research has shown that traditional fit indices [e.g., root mean square error of approximation ($RMSEA$), comparative fit index (CFI), Tucker–Lewis Index ($TLLI$), and standardized root mean square residual ($SRMR$)] are not sensitive to misspecified between-level models in multilevel structural equation modeling (MSEM) (Hsu et al., 2017; Padgett & Morgan, 2020). For this reason, researchers have advocated the use of level-specific (l-s) fit indices for evaluating the within-level model and the between-level model separately (Hox, 2010; Hsu et al., 2017; Rappaport et al., 2020; Ryu, 2014; Ryu & West, 2009; Schermelleh-Engel et al., 2014; Wu

✉ John J. H. Lin
john.jrhunglin@ntnu.edu.tw

¹ Graduate Institute of Science Education, National Taiwan Normal University, No. 88, Ting-Jou Rd., sec. 4, Taipei, Taiwan

² Department of Curriculum and Instruction, University of Massachusetts Lowell, Lowell, MA, USA

et al., 2017). According to Ryu and West (2009), the 1-s fit indices could be straightforwardly computed by the partially saturated-model method (PS method). Note that the existing guidelines for using level-specific fit indices are based on population multilevel measurement models with continuous indicators, and few address ordered categorical variables (Padgett & Morgan, 2020). Yet it is unclear whether these guidelines could be applicable to models with dichotomous indicators.

To shed light on this issue, this study endeavors to examine the performance of level-specific χ^2 test statistics and fit indices derived from the PS method in terms of their sensitivity to lack of fit at specific levels in MCFA with dichotomous indicators. In addition, the effectiveness of alternative level-specific fit indices obtained from Mplus—SRMR for the within-level model ($SRMR_w$) and for the between-level model ($SRMR_B$)—was compared with PS-level-specific fit indices.

MCFA with dichotomous indicators in multilevel structural equation modeling

For simplicity, we consider a two-level single-factor model. Let Y_{pig} denote the p^{th} dichotomous indicator (i.e., latent variable indicator) for individual i nested within g group ($p = 1 \dots P$ dichotomous indicators, $i = 1 \dots N$ individuals, and $g = 1 \dots G$ groups),

$$Y_{pig} = \begin{cases} 1, & \text{if } y_{pig}^* > \tau \\ 0, & \text{if } y_{pig}^* \leq \tau \end{cases} \tag{1}$$

The equation expresses a threshold model which assumes that underlying the dichotomous indicator Y_{pig} is a normally distributed continuous latent variable y_{pig}^* , which can determine the category of dichotomous indicator by the threshold (τ) (Asparouhov & Muthen, 2007; Bollen, 2002). That is, the indicators of interest are conceptualized as continuous, but the format of response to each indicator is in a restrictive, dichotomous scale (Bollen, 2002). For example, if the i^{th} individual falls short of the threshold, the response of this individual would be 0. If the i^{th} individual passes the threshold, the response of this individual would be 1.

Using similar notations to those used by Padgett and Morgan (2021), in this section, we outline a two-level measurement model with dichotomous indicators, as illustrated in Fig. 1. Note that the model in Fig. 1 was also adopted as a population model for simulated dataset generation in the current study. Using the between-and-within specification approach (Hox, 2010; Muthen, 1994), the covariance structure is partitioned into a within-level component (denoted by a W subscript) and a between-level component (denoted by a B subscript). Separate models are specified for each component. The within-level component captures individual-level variation, while the between-level component captures variation between groups.

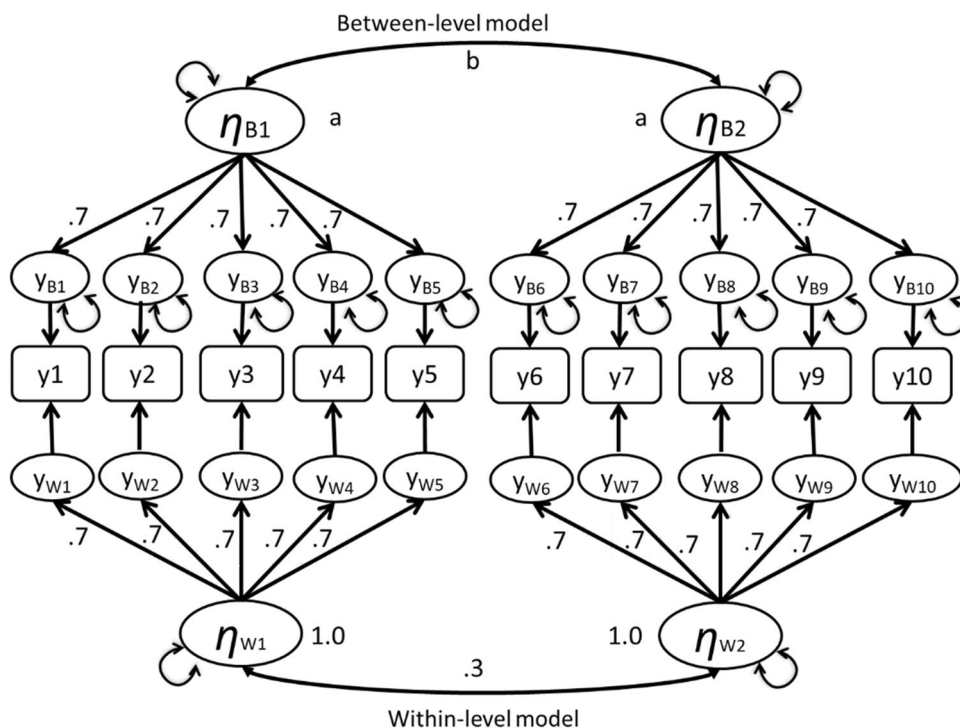


Fig. 1 A two-level data generating model with two factors at each level

Let y_{ig} denote the p -dimensional response vector for student i in group g . The response vector y_{ig} is decomposed as:

$$y_{ig} = \mu + y_{w_{ig}} + y_{B_g} \quad (2)$$

where μ represents the grand mean, and $y_{w_{ig}}$ and y_{B_g} are independent within-level and between-level components, respectively. The measurement model at the within-level is given by the equation:

$$y_{ig} = \mu_g + \Lambda_W \eta_{W_{ig}} + \epsilon_{W_{ig}} \quad (3)$$

where Λ_W is the $p \times 2$ factor loadings matrix for the within-level latent factor ($\eta_{W_{ig}}$), $\eta_{W_{ig}}$ vector is the distributed multivariate normal with an expectation of zero and is a 2×2 covariance matrix Ψ_W . The $\epsilon_{W_{ig}}$ is multivariate normal distributed with an expectation of zero and $p \times p$ diagonal covariance matrix Θ_W , with error terms along the diagonal.

The measurement model at the between-level is given by

$$\mu_g = \mu + \Lambda_B \eta_{B_g} + \epsilon_{B_g} \quad (4)$$

Here, Λ_B , η_{B_g} , and ϵ_{B_g} are the between-level terms corresponding to the within-level terms Λ_W , $\eta_{W_{ig}}$, and $\epsilon_{W_{ig}}$. Moreover, the covariance matrices Ψ_B and Θ_B are the between-level counterparts to the within-level covariance matrices Ψ_W and Θ_W . We can obtain Eq. (5) after combining Eqs. (3) and (4):

$$y_{ig} = \mu + \Lambda_W \eta_{W_{ig}} + \Lambda_B \eta_{B_g} + \epsilon_{W_{ig}} + \epsilon_{B_g} \quad (5)$$

Estimator

The diagonally weighted least squares (DWLS) estimator is recommended for single-level confirmatory factor analysis (CFA) with categorical indicators (DiStefano & Morgan, 2014; Forero et al., 2009) and multilevel CFA due to its ability to identify the correct model specification (Padgett & Morgan, 2020). The DWLS estimator was based on polychoric correlation and the inverse of the asymptotic covariance matrix W^{-1} of the sample variances and covariances as a weight matrix. Because the estimation of W^{-1} is quite unstable when the sample size is small, the DWLS uses only the diagonal elements of W in model fitting and uses full W to obtain standard errors and χ^2 values (Jöreskog & Sörbom, 1996). As a result, the DWLS produces robust standard errors and χ^2 values (Flora & Curran, 2004; Muthen, 1993). Finney et al., (2006) suggested that when fewer than five categories are used, the DWLS estimator resulted in robust parameter estimates, standard errors, and fit indices for models with categorical nature of the data. In addition, the DWLS estimator has been found to perform well with small sample sizes and large models (Flora & Curran,

2004; Yang-Wallentin et al., 2010). Beauducél and Herzberg (2006) found that DWLS produced fit indices (root mean square error of approximation [*RMSEA*], comparative fit index [*CFI*], Tucker-Lewis Index [*TLI*]) that adequately indicated correctly specified models. However, when data were non-normally distributed in two to four categories, Bandalos (2008) found that robust DWLS-based *RMSEA* and *CFI* inadequately identified poorly misspecified models. Note that the DWLS estimator is known as weighted least squares mean and variance adjusted (WLSMV) estimator in Mplus. In the present study, the DWLS estimator was adopted for analyzing simulated datasets by using the command “estimator = WLSMV” in Mplus.

Cut-off values for using fit indices on CFA with dichotomous indicators

Several simulation studies have examined whether Hu and Bentler’s (1999) conventional cut-off values for traditional fit indices (i.e., *RMSEA* < .06; *CFI* and *TLI* > .95; *SRMR* < .08) can be applied similarly when DWLS is used to ordered categorical data. In general, prior studies suggested that conventional cut-off values should be applied with careful consideration of the data’s characteristics, such as sample size, asymmetry of categorical data, and number of categories (DiStefano & Morgan, 2014; Garrido, Abad, & Ponsoda, 2016; Nye & Drasgow, 2011; Xia & Yang, 2019). To the best of our knowledge, Padgett and Morgan (2021) is the only study that provided recommended cut-off criteria for using traditional fit indices in MCFA with categorical indicators. Specifically, Padgett and Morgan found that *CFI*, *TLI*, and *RMSEA* were primarily influenced by within-level misspecification, but partially influenced by between-level misspecification. In addition, the performance of the three fit indices was impacted by the sample size at the between-level (N_2) when DWLS was used. Although they provide recommended cut-off criteria for *CFI* and *TLI* (> .98 if $N_2 < 100$; > .97 if $N_2 \geq 100$), and *RMSEA* (< .02 regardless of N_2), they also cautioned that those fit indices should be used only to provide weak evidence for some type of misspecification. Alternatively, *SRMR_w* and *SRMR_B* may provide within-level and between-level model-data fit, respectively. When DWLS is used, *SRMR_w* needs a cut-off value of < .05 if $N_2 < 100$, and a cut-off value of < .04 if $N_2 \geq 100$. On the other hand, *SRMR_B* is suggested to have a cut-off value of < .06 if $N_2 \geq 100$ and should not be used if $N_2 < 100$.

Model evaluation: PS-level-specific fit indices

Numerous simulation studies have indicated that level-specific tests of exact fit (i.e., χ^2 test statistics) and fit indices derived from the partially saturated-model method were recommended for detecting misspecification in MCFA (Hsu

et al., 2017; Lee & Sohn, 2022; Rappaport et al., 2020; Ryu, 2011, 2014; Ryu & West, 2009; Schermelleh-Engel et al., 2014). Using the PS method, researchers can first derive the between-level-specific (b-l-s) χ^2 test statistics ($\chi^2_{PS_B}$). Specifically, $\chi^2_{PS_B}$ can be derived by specifying a hypothesized between-level model and saturating the within-level model (i.e., correlating all observed variables). A saturated within-level models can be seen as a just-identified model with zero degrees of freedom, and thus has a χ^2 test statistic equal to zero. Consequently, $\chi^2_{PS_B}$ only reflects the model fit of the hypothesized between-level model. Since fit indices are a function of χ^2 test statistics, researchers can then compute b-l-s fit indices ($RMSEA_{PS_B}$, CFI_{PS_B} , TLI_{PS_B}) using the value of $\chi^2_{PS_B}$.

In the same manner, within-level-specific (w-l-s) χ^2 test statistics ($\chi^2_{PS_W}$) can be derived by specifying a hypothesized within-level model and saturating the between-level model. The w-l-s fit indices ($RMSEA_{PS_W}$, CFI_{PS_W} , TLI_{PS_W}) can be computed using the value of $\chi^2_{PS_W}$. The formulas for computing l-s fit indices are identical to those in Ryu and West (2009) and Hsu, Lin, Skidmore, and Kim (2018) studies (see Appendix A). Additionally, the performance of the aforementioned l-s fit indices was compared with that of two alternative l-s fit indices, $SRMR_W$ and $SRMR_B$, which are computed based on the discrepancy between the sample covariance and the corresponding model-implied covariance. Both $SRMR_W$ and $SRMR_B$ are available in the Mplus model solution output.

Previous simulation studies on L-s fit indices

Prior simulation studies have investigated the performance of l-s fit indices when the multivariate normality assumption was met in MCFA. Ryu and West (2009) examined the effectiveness of b-l-s fit indices ($RMSEA_{PS_B}$, CFI_{PS_B}) and w-l-s fit indices ($RMSEA_{PS_W}$, CFI_{PS_W}) using a population MCFA model with continuous indicators where multivariate normality was assumed. The intraclass correlation coefficient (ICC) level in the population model was fixed to 0.5. Ryu and West (2009) found that both types of fit indices can correctly indicate good and poor model fit in various sample size conditions (numbers of groups = 50, 100, 200, and 1000, and group size = 20, 50, 100). Ryu and West (2009), however, also found that a sample size of 50 groups could cause nonconvergence problems. Ryu and West's findings were validated by Rappaport et al. (2020).

Hsu et al. (2017) extended the findings of Ryu and West (2009) by investigating the performance of l-s fit indices when ICC was smaller than 0.5. Hsu et al. (2017) found that the performance of w-l-s fit indices ($RMSEA_{PS_W}$, CFI_{PS_W} , TLI_{PS_W} and $SRMR_W$) was barely influenced by ICC, while b-l-s fit indices ($RMSEA_{PS_B}$, CFI_{PS_B} , TLI_{PS_B} and $SRMR_B$) were less-promising indicators to correctly indicate good

or poor model fit. Similarly, Lee and Sohn (2022) found that w-l-s fit indices were sensitive to detecting misspecified within-level models and were less impacted by ICC and sample size. In addition, Lee and Sohn discovered that $RMSEA_{PS_B}$ and $SRMR_B$ were more promising for detecting misspecified between-level models with an increase in ICC, while CFI_{PS_B} and TLI_{PS_B} were also influenced by ICC, but the influence was moderated by the type of misspecifications in models (e.g., misspecification in factor cross-loadings or factor covariance). In summary, previous studies have considered various design factors, such as numbers of groups, group size, ICC, and type of misspecifications.

To the best of our knowledge, the performance of l-s fit indices in MCFA with categorical indicators has not been well examined in prior research. Hsu (2009) examined the sensitivity of $SRMR_W$ and $SRMR_B$ in MCFA with dichotomous indicators. Hsu (2009) found that $SRMR_W$ can correctly indicate good model fit (i.e., type I error rate < .05) and poor model fit (i.e., statistical power > .80) due to an intentional misspecification in factor covariance. The type I error rate of $SRMR_B$, however, tended to be higher across different conditions, and the statistical power was less satisfying when ICC was low. Similar findings about $SRMR_W$ and $SRMR_B$ were revealed in Navruz's (2016) study.

The present study

To date, no studies have attempted to evaluate the performance of l-s fit indices derived from the PS method in MCFA with categorical indicators. To address this concern, the present study attempted to evaluate the performance of commonly used l-s fit indices using a population model with dichotomous indicators. The findings of this study may shed some light on the practice of model evaluation when categorical indicators are used in MCFA.

The design factors considered in the present study included numbers of groups, group size, ICC, as well as two other factors, thresholds of dichotomous indicators and factor loadings, which have not been widely investigated in previous simulation studies focusing on effectiveness of l-s fit indices. Asymmetric thresholds of dichotomous indicators could occur in real-world situations, and prior studies have discovered the impact of asymmetric thresholds on the performance of several scaled or adjusted χ^2 statistics (hereafter called adjusted χ^2 statistics). For example, Rhemtulla, Brosseau-Liard, and Savalei (2012) examined three different conditions of thresholds (50:50, 60:40, and 80:20) and found that adjusted χ^2 statistics had relatively low statistical power in detecting serious model misspecification with asymmetric thresholds and small samples. In addition, type I error rates were reasonable (smaller than .05) when the threshold was symmetric, but extreme asymmetry thresholds (80:20) could cause high type I error

rates. Savalei and Rhemtulla (2013) conducted simulations based on a two-factor model with dichotomous indicators using a DWLS estimator. They examined the difference between symmetric thresholds (50:50) and asymmetric thresholds (64:36 and 85:15) on adjusted χ^2 statistics examined in Rhemtulla et al. (2012). Results suggested that the performance of adjusted χ^2 statistics decreases as thresholds become more asymmetric. To the best of our knowledge, the impact of asymmetric thresholds on l-s fit indices has not been well documented. Because fit indices were a function of χ^2 statistics, the performance of fit indices could very likely be influenced by the thresholds of dichotomous indicators. For this reason, we examined the impact of asymmetric thresholds on level-specific fit indices in MCFA with dichotomous indicators.

In addition, previous simulation studies (e.g., Forero et al., 2009; Garrido et al., 2016; Heene et al., 2011; Nestler, 2013) have consistently shown that CFAs with low factor loadings resulted in less adequate model evaluation results. Nestler (2013) examined the impact of factor loadings—high (0.70), medium (0.55), and low (0.40)—on χ^2 statistics with a two-factor CFA with dichotomous indicators. Nestler found that when factor loadings were high, the statistical power of χ^2 test statistics was satisfying ($\sim .80$), regardless of sample size. However, when factor loadings were medium or low, a sample size of 250 or 500, respectively, was needed to retain satisfactory statistical power. Furthermore, previous simulation studies have shown that the magnitudes of factor loadings had an impact on the performance of traditional fit indices. For example, Heene et al. (2011) examined how the performance of the χ^2 test statistic, *RMSEA*, *CFI*, and *SRMR* could be influenced by three levels of factor loadings [low $\sim (0.30, 0.50)$; medium $\sim (0.50, 0.70)$; high $\sim (0.70, 0.90)$]. Heene et al., 2011 found that decreasing factor loadings led to decreasing values of the χ^2 test statistic and three fit indices, altering the statistical power to detect misspecified models. In general, low factor loadings resulted in decreasing values of χ^2 test statistic and *RMSEA*, *CFI*, and *SRMR*. As a result, misspecification was often not detected by the χ^2 test statistic, *RMSEA*, or *SRMR* when factor loadings were low. However, high factor loadings can cause the rejection of just slightly misspecified models (cf. Savalei, 2012). In contrast, Heene et al., 2011 found that the *CFI* tended to exhibit poorer fit for models with low factor loadings. The reason is that models with lower factor loadings held lower covariances between the observed variables. As a result, the distance between the hypothesized model and the baseline null model would be reduced, resulting in lower *CFI* values (Garrido et al., 2016). The impact of low factor loadings on *CFI* can be extended to *TLI* because both fit indices are a function of the distance between the hypothesized model and the baseline null model. To date, few efforts have been made to study the impact of factor loadings on l-s fit indices

in MCFA with dichotomous indicators. Our study aimed to bridge this knowledge gap.

Note that although we were aware that type of misspecification (MT) can be manipulated as a design factor, our study only considered one MT condition, which was the misspecification in factor covariance. Our selected misspecification scenario is important for applied researchers who wish to verify the construct validity of their instruments. We did not include the scenario of misspecification in cross-loadings as another MT condition because we think misspecification in cross-loadings is a highly complex topic (e.g., over-specification, under-specification, magnitudes of factor cross-loadings) which deserves a new study to comprehensively investigate this issue.

Method

A Monte Carlo study was conducted to evaluate the performance of l-s fit indices in MCFA with dichotomous indicators. The five design factors examined in this study were numbers of groups, group size, intraclass correlation coefficient, thresholds of dichotomous indicators, and factor loadings.

Population model

In the current study, the population model (see Fig. 1) for simulation data generation was based on the population model presented in Hsu et al.'s (2017) study. The population model was a two-level measurement model with two within-level factors (η_{W1} and η_{W2}) and two between-level factors (η_{B1} and η_{B2}). At the within-level, five dichotomous observed indicators were loaded on each factor. Parameters in the within-level model for simulation data generation were factor loadings = 0.70, factor variances = 1.00, and factor covariance = 0.30. The residual variance parameters cannot be freely estimated; therefore, no initial residual variances were set (Muthen, 1990). Factors and residual variances were independent of each other. Note that the correlation between two within-level factors was also 0.30 because within-level factor variances were fixed at 1.00. The threshold of indicators was equal to 0, resulting in a 50:50 proportion of responses that are 0 or 1.

The between-level model had an identical factorial structure to the within-level model. Parameters for simulation data generation were factor loadings = 0.70, residual variances = 0.51. Factors and residuals were independent of each other. We varied the variance of between-level factors (a in Fig. 1) to create three ICC conditions. (More detailed information is presented in the next section.) Note that for each ICC condition, we adjusted the between-level factor covariance (b in Fig. 1) based on the formula: 0.30

$\times \sqrt{\text{var}(\eta_{B1})} \times \sqrt{\text{var}(\eta_{B2})}$ so that the correlation of two between-level factors can be held to 0.30 across different ICC conditions.

Design factors

Numbers of groups (NG) A NG larger than 100 was recommended as an acceptable estimate at the between-level with low ICC (Hox & Maas, 2001; Hsu et al., 2015). Ryu and West (2009) used NG = 50, 100, 200, and 1000 in their simulation work, where NG = 50 could lead to nonconvergence problems and NG = 1000 was not a realistic NC for practitioners. Hsu et al. (2017) found that NG = 200 seemed to compensate for convergence problems if ICC was low. As a result, this study considered NG ranging from 50 to 200. Specifically, the current study adopted three NG conditions (50, 100, and 200) to evaluate the impact of NC on the performance of level-specific fit indices when the indicators of MCFA were dichotomous.

Group size (GS) Hox and Maas (2001) used a set of GS = 10, 20, 50 on a MSEM study and found that GS had a trivial impact on parameter estimates, as well as standard errors. Additionally, Ryu and West (2009) used GS = 20, 50, and 100 in their study focusing on performance of fit indices of MSEM. In consideration of common practices and comparability, the present study adopted GS = 10, 20, and 50.

ICC. The ICC (ρ) is defined as:

$$\rho = \frac{VAR_{Between}}{VAR_{Between} + VAR_{Within}} \quad (6)$$

where $VAR_{Between}$ and VAR_{Within} are the variance of between-level factors and within-level factors, respectively. The variance of within-level factors was constrained to 1.00, while that of between-level factors varied to create different ICC conditions. The present study considered three levels of between-level variance (0.1, 0.3, and 0.5), resulting in three levels of ICC: .091 (ICC1), .231 (ICC2), and .333 (ICC3).

Thresholds of dichotomous indicators (THR) To the best of our knowledge, study of the impact of asymmetric thresholds on 1-s fit indices is still lacking. Guided by Rhemtulla et al.'s (2012) and Savalei and Rhemtulla's (2013) studies, we considered both symmetric (50:50) and asymmetric (80:20) conditions. By considering these design factors, we intended to understand the impact of asymmetry of categorical data on the performance of 1-s χ^2 and fit indices.

Factor loadings (FL) Following Nestler (2013) simulation study, we adopted three conditions of FL in this study: low (.40), medium (.55), and high (.70). The setting of FL was also in line with previous simulation studies (Forero et al.,

2009; Garrido et al., 2016; Heene et al., 2011). Previous simulation studies have shown that low magnitudes of factor loadings resulted in less sensitivity of traditional fit indices. By including this design factor, we intended to examine the extent to which the magnitudes of factor loadings can impact the performance of 1-s χ^2 and fit indices.

As a result, a total of 162 conditions (NG = 50, 100, and 200; GS = 10, 20, and 50; ICC = ICC1 to ICC3; THR = 50:50 and 80:20; FL = .40, .55, and .70) were yielded. For each condition, 500 replications were generated using Mplus 7.4 (Muthen & Muthen, 1998–2015).

Intentional misspecifications in the hypothesized models

After simulation data were generated, we analyzed simulation data in three different conditions: (a) correctly specified hypothesized model (i.e., the hypothesized model was equal to the population model as shown in Fig. 1, M_C); (b) misspecification in between-level model only (M_B , see Fig. 2); and (c) misspecification in within-level model only (M_W , see Fig. 2). Following Ryu and West's (2009) and Hsu et al.'s (2017) studies, M_B contained a misspecification where only one between-level factor loaded on the indicators. M_W contained a misspecification where only one within-level factor loaded on the indicators. The WLSMV estimator was applied in three conditions to obtain the model solutions in Mplus. Starting values of parameter estimates were set to the same as the parameter values in the population model to prevent any convergence problems due to bad starting values. Fit indices of interest were computed in each condition.

Analysis

In M_C , M_B , and M_W conditions, the values of fit indices were saved for subsequent analyses. Deceptive statistics for level-specific χ^2 test statistics and fit indices were computed and reported. If needed, factorial analysis of variance (ANOVA) was applied to determine the impact of the design factors on the performance of fit indices (Skrondal, 2000). Specifically, eta-squared (η^2) was reported to indicate the proportion of the variance accounted for by a particular design factor or the interaction effect terms. Following Cohen's (1988, 1992) suggestion, we used a moderate η^2 above .06 (i.e., practically significant) to identify influential design factors in the fit indices values. Note that when a fit index had a standard deviation close to 0, the impact of design factors on the values of the fit index were self-evidently trivial, regardless of the η^2 values. In addition, for level-specific χ^2 test statistics, we computed the type I error rates (under M_C) or statistical power (under M_B and M_W) under α level = .05.

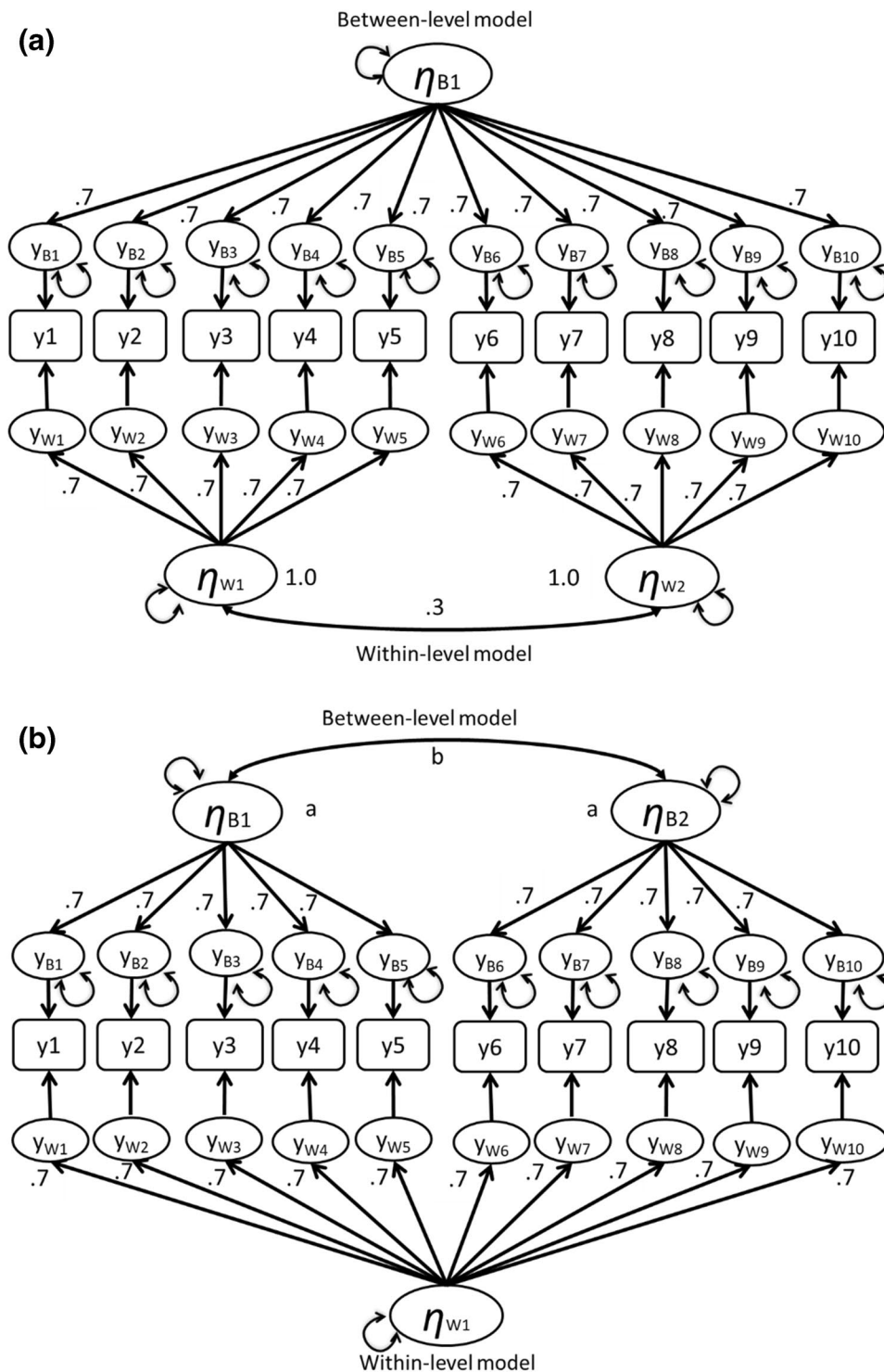


Fig. 2 **a** Misspecification in the between-level model only (M_B). **b** Misspecification in the within-level model only (M_W)

For level-specific fit indices, we applied the traditional cut-off values ($RMSEA$ -related fit indices $< .06$; CFI - and TLI -related fit indices $> .95$; $SRMR$ -related fit indices $< .08$; Hu & Bentler, 1999) to explore whether these fit indices were promising to indicate correctly specified or misspecified

hypothesized models. Note that our intention was not to encourage using traditional cut-off values as golden rules for model evaluation. Rather, we intended to provide a sense of whether these cut-off values are applicable when level-specific χ^2 test statistics or fit indices are used.

Results

Convergence rates

The convergence rates were highly associated with magnitude of ICC and FL, followed by sample size (NG and GS). In contrast, THR was less influential. In this section, the convergence rates were reported by ICC and FL. First, the convergence rates were above 95% in conditions with high ICC (.333) and high FL (.70), with one exception (the convergence rate was 92% when NG = 50, GS = 10, and THR = 80:20). In conditions with high ICC and medium FL (.55), the coverage rates were close to or above 95% only when NG/GS were at least 100/20 or NG/GS were at least 200/10. On the other hand, in conditions with high ICC and low FL (.40), the coverage rates were below 95% (ranging between 46% and 89%). Second, in conditions with medium ICC (.231) and high FL, the convergence rates were above 95% when NG/GS were at least 100/10 and THR was symmetric. In contrast, in conditions with medium ICC and medium FL, the convergence rates were above 95% only when NG/GS were at least 200/50. Unfortunately, in conditions with medium ICC and low FL, the coverage rates were below 95% (ranging between 37% and 63%). Third, in conditions with low ICC (.091), the coverage rates were below 95% – in range of 48% and 84% when FL was high; in range of 40% and 52% when FL was medium; and in range of 29% and 47% when FL was low. Only converged solutions were used for further analysis.

Effects of design factors on level-specific fit indices

The left-hand side of Table 1 summarizes the descriptive statistics [aggregated means and standard deviations (SDs)] of level-specific χ^2 test statistics and fit indices for the M_C , M_B , and M_W conditions. The right-hand side of Table 1 shows the η^2 values derived from ANOVA results using level-specific χ^2 test statistics and fit indices as outcomes. We have highlighted η^2 values above .06 (Cohen, 1988, 1992) in grey to show practically significant effects of design factors. Note that when a level-specific χ^2 test statistic or fit index has a small variation (indicated by a small SD), the impact of design factors was self-evidently trivial even if any η^2 exceeding .06 is identified. Note the η^2 values of three-way interactions were close to 0, and therefore, were not reported in Table 1 for the sake of simplicity. In addition, to inform the performance of level-specific χ^2 test statistics and fit indices, we report the average values of level-specific χ^2 test statistics and fit indices by all design factors in Appendix B (Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21), where type I error rates or statistical power are reported for level-specific χ^2 test statistics.

In the following section, we reported the simulation of between-level-specific and within-level-specific χ^2 and fit indices separately. To better inform interested readers of our findings, we created Tables 2 and 3 to summarize the simulation results for between-level-specific and within-level-specific χ^2 and fit indices, respectively. In addition, our simulation results suggest the performance of between-level-specific χ^2 and fit indices was complicated and more likely influenced by NG, ICC, and FL. We visualized the values of between-level-specific fit indices in Figs. 3 and 4 (by NG, ICC, and FL with GS set to 50.) for M_C and M_B conditions, respectively. Also, because we found THR could slightly weigh in the performance of between-level-specific fit indices in some conditions, the visualization was presented for symmetric THR and asymmetric THR conditions separately. Note we did not present the visualization of within-level-specific fit indices because the patterns of their performance were less complicated and could be clearly described in Table 3.

Between-level-specific χ^2 and fit indices

χ_B^2 As shown in Table 1, under M_C where the hypothesized between-level model was correctly specified, χ_B^2 had a mean of 31.53, an SD of 6.45, and the η^2 s ranged from .00 to .03. Small values of η^2 s suggest that five design factors (NG, GS, ICC, THR, and FL) accounted for trivial variance of χ_B^2 . That is, our design factors had a negligible effect on χ_B^2 . Appendix B shows that the type I error rates of χ_B^2 were generally acceptable (close to or below .05) across all conditions. On the other hand, under M_B , where the hypothesized between-level model was misspecified, χ_B^2 had a mean of 45.22 and an SD of 23.77 (see Table 1). FL ($\eta^2 = .10$), ICC ($\eta^2 = .09$), NG ($\eta^2 = .06$), and ICC*FL ($\eta^2 = .07$) can jointly affect χ_B^2 . Based on data in Appendix B, we found χ_B^2 had higher statistical power to detect misspecified between-level models when FL, ICC, or NG increased. These results show that when χ_B^2 was used, type I error rates might not be a concern, but statistical power would be of problem when FL, ICC, or NG were low.

We further explore the statistical power of χ_B^2 reported in Appendix B. First, we found statistical power was far below .80 under several conditions: NG = 50 or GS = 10, or ICC1 or low FL. Second, we found χ_B^2 was able to reach satisfying statistical power (close to or above .80) under ICC3 with (a) high FL and NG = 100, or (b) medium FL and NG = 200. If ICC was medium (ICC2), only the following condition resulted in satisfying statistical power: high FL, NG/GS at least 200/20, and symmetric THR. The findings regarding the performance of χ_B^2 were summarized in Table 2.

RMSEA_B Table 1 shows that under M_C , $RMSEA_B$ had a close-to-zero mean (0.01) and SD (0.02), suggesting that $RMSEA_B$

Table 1 ANOVA results (η^2) with χ^2 test statistics and fit indices values as the dependent variables for the * NG * GS * ICC * THR * FL design

ANOVA Results (η^2)									
<i>M_C</i>	<i>B-I-s</i>	Mean (<i>SD</i>)	NG	GS	ICC	THR	FL	NG*GS	NG*ICC
	$\chi^2_B(df=34)$	31.53(6.45)	0.02	0.01	0.02	0.01	0.03	0.01	0.00
	<i>RMSEA_B</i>	0.01(0.02)	0.00	0.01	0.01	0.01	0.02	0.00	0.00
	<i>CFI_B</i>	0.81(0.35)	0.08	0.01	0.09	0.01	0.10	0.00	0.00
	<i>TLI_B</i>	0.79(0.36)	0.07	0.01	0.09	0.01	0.09	0.00	0.00
	<i>SRMR_B</i>	0.09(0.03)	0.56	0.09	0.00	0.02	0.00	0.01	0.00
	W-I-s								
	$\chi^2_W(df=34)$	31.84(7.14)	0.01	0.09	0.00	0.00	0.01	0.01	0.00
	<i>RMSEA_w</i>	0.00(0.01)	0.01	0.10 ^a	0.00	0.00	0.01	0.01	0.00
	<i>CFI_w</i>	0.99(0.03)	0.02	0.07	0.00	0.01	0.03	0.02	0.00
	<i>TLI_w</i>	0.99(0.03)	0.02	0.08	0.00	0.01	0.03	0.02	0.00
	<i>SRMR_w</i>	0.04(0.02)	0.29	0.40	0.00	0.08	0.01	0.03	0.00
<i>M_B</i>	B-I-s								
	$\chi^2_B(df=35)$	45.22(23.77)	0.06	0.00	0.09	0.01	0.10	0.01	0.04
	<i>RMSEA_B</i>	0.03(0.04)	0.02	0.01	0.14	0.01	0.14	0.00	0.01
	<i>CFI_B</i>	0.65(0.34)	0.04	0.00	0.02	0.00	0.03	0.00	0.01
	<i>TLI_B</i>	0.59(0.36)	0.03	0.00	0.01	0.00	0.01	0.00	0.01
	<i>SRMR_B</i>	0.10(0.03)	0.42	0.07	0.03	0.01	0.08	0.01	0.01
<i>M_W</i>	W-I-s								
	$\chi^2_W(df=35)$	306.75(377.37)	0.11	0.08	0.00	0.03	0.14	0.05	0.00
	<i>RMSEA_w</i>	0.04(0.02)	0.03	0.01	0.00	0.09	0.56	0.00	0.00
	<i>CFI_w</i>	0.74(0.08)	0.04	0.01	0.00	0.02	0.09	0.00	0.00
	<i>TLI_w</i>	0.66(0.10)	0.04	0.01	0.00	0.02	0.09	0.00	0.00
	<i>SRMR_w</i>	0.09(0.02)	0.05	0.06	0.01	0.00	0.54	0.01	0.00
ANOVA Results (η^2)									
<i>M_C</i>	<i>B-I-s</i>	NG*THR	NG*FL	GS*ICC	GS*THR	GS*FL	ICC*THR	ICC*FL	THR*FL
	$\chi^2_B(df=34)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	<i>RMSEA_B</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	<i>CFI_B</i>	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
	<i>TLI_B</i>	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
	<i>SRMR_B</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	W-I-s								
	$\chi^2_W(df=34)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	<i>RMSEA_w</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	<i>CFI_w</i>	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00
	<i>TLI_w</i>	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00
	<i>SRMR_w</i>	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00
<i>M_B</i>	B-I-s								
	$\chi^2_B(df=35)$	0.00	0.04	0.00	0.00	0.01	0.00	0.07	0.00
	<i>RMSEA_B</i>	0.00	0.01	0.00	0.00	0.01	0.00	0.08	0.01
	<i>CFI_B</i>	0.00	0.01	0.00	0.00	0.01	0.00	0.02	0.00
	<i>TLI_B</i>	0.00	0.01	0.00	0.00	0.01	0.00	0.03	0.00
	<i>SRMR_B</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
<i>M_W</i>	W-I-s								
	$\chi^2_W(df=35)$	0.01	0.06	0.00	0.01	0.06	0.00	0.00	0.02
	<i>RMSEA_w</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03
	<i>CFI_w</i>	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	<i>TLI_w</i>	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	<i>SRMR_w</i>	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01

Note. *M_C* = correctly specified hypothesized model. *M_B* = misspecification in between-level model only. *M_W* = misspecification in within-level model only. B-I-s = between-level specific. W-I-s = within-level specific. NG = numbers of groups. GS = group size. ICC = intraclass correlation coefficient. THR = threshold of dichotomous indicators. FL = factor loadings. Note that we used a moderate η^2 above .06 (highlighted in gray) to identify influential design factors. ^a η^2 is less practically meaningful because the fit index demonstrated extremely low variability. The η^2 values of three-way interactions were close to 0, and therefore, were not reported for the sake of simplicity.

Table 2 Summary of simulation findings for between-level-specific χ^2 and fit indices

Between-level-specific χ^2 and fit indices	Findings
χ_B^2	<p>M_C: Design factors had no impacts.</p> <p>M_B: ICC and FL had a similar magnitude of impacts on χ_B^2; NG and ICC*FL had relatively smaller impacts. χ_B^2 had stronger statistical power to correctly indicate poor models when ICC, FL, or NG increased.</p> <p>Findings regarding type I error rates and statistical power:</p> <ul style="list-style-type: none"> • Type I error rates of χ_B^2 were close to or below .05 across all conditions. • Statistical power of χ_B^2 was far below .80 under several conditions: NG=50, GS=10, ICC1, or low FL. • Statistical power of χ_B^2 was close to or above .80 in following conditions: <ul style="list-style-type: none"> ○ When ICC was high (ICC3), high FL with NG=100 or medium FL with NG=200 can reach satisfying statistical power. ○ When ICC was medium (ICC2), only the following condition resulted in satisfying statistical power: high FL, NG/GS at least 200/20, and symmetric THR.
$RMSEA_B$	<p>M_C: Design factors had no impacts.</p> <p>M_B: ICC, FL, and ICC*FL had impacts on $RMSEA_B$. $RMSEA_B$ was more promising to correctly indicate poor models when ICC or FL increased.</p> <p>Findings if the cut-off value of .06 was applied:</p> <ul style="list-style-type: none"> • $RMSEA_B$ cannot be used to determine model fit under ICC1 or low FL conditions. • $RMSEA_B$ can be used to determine model fit under following conditions: <ul style="list-style-type: none"> ○ When ICC was high (ICC3) with high FL, $RMSEA_B$ was useful except for one condition: asymmetric THR and NG/GS=50/10. ○ When ICC was high with medium FL, $RMSEA_B$ was useful if NG/GS at least 200/20. ○ When ICC was medium (ICC2), $RMSEA_B$ performed well only if high FL, symmetric THR, and NG/GS at least 100/20.
CFI_B	<p>M_C: NG, ICC, and FL had a similar magnitude of impacts; CFI_B was more promising to identify correctly hypothesized between-level models when NG, ICC, or FL increased.</p> <p>M_B: Design factors had no impacts.</p> <p>Findings if the cut-off value of .95 was applied:</p> <ul style="list-style-type: none"> • CFI_B cannot be used to determine model fit under ICC1 or low FL conditions. • CFI_B can be used to determine model fit under following conditions: <ul style="list-style-type: none"> ○ When ICC was high (ICC3) with high FL, CFI_B was useful if NG/GS at least 50/20 or NG/GS at least 100/10. ○ When ICC was high with medium FL, CFI_B was useful if NG/GS at least 100/20 or NG/GS at least 200/10. ○ When ICC was medium (ICC2) with high FL, CFI_B was useful (a) if NG/GS at least 100/50 or NG/GS at least 200/10 with symmetric THR; or (b) if NG/GS at least 200/50 with asymmetric THR. ○ When ICC was medium with medium FL, CFI_B was useful if NG/GS at least 200/50.
TLI_B	<p>M_C: NG, ICC, and FL had a similar magnitude of impacts; TLI_B was more promising to identify correctly hypothesized between-level models when NG, ICC, or FL increased.</p> <p>M_B: Design factors had no impacts.</p> <p>Findings if the cut-off value of .95 was applied:</p> <ul style="list-style-type: none"> • TLI_B cannot be used to determine model fit under ICC1 or low FL conditions. • TLI_B can be used to determine model fit under following conditions: <ul style="list-style-type: none"> ○ When ICC was high (ICC3) with high FL, TLI_B was useful if NG/GS = 50/50 or NG/GS at least 100/10. ○ When ICC was high with medium FL, TLI_B was useful if NG/GS at least 100/50 or NG/GS at least 200/10. ○ When ICC was medium (ICC2) with high FL, TLI_B was useful (a) if NG/GS at least 100/50 or NG/GS at least 200/10 with symmetric THR; or (b) if NG/GS at least 200/50 with asymmetric THR. ○ When ICC was medium (ICC2) with medium FL, TLI_B was useful if NG/GS at least 200/50.
$SRMR_B$	<p>M_C: NG had major impacts on $SRMR_B$, followed by GS. $SRMR_B$ was more promising to identify correctly hypothesized between-level models when NG or GS increased.</p> <p>M_B: NG had major impacts on $SRMR_B$, followed by GS and FL, which had a similar magnitude of impacts. $SRMR_B$ was less promising to correctly indicate poor models when NG or GS increased; $SRMR_B$ was more promising when FL increased. ICC was a nonignorable factor when cut-off value of .08 was used.</p> <p>Findings if the cut-off value of .08 was applied:</p> <ul style="list-style-type: none"> • $SRMR_B$ cannot be used to determine model fit under (a) low NG, (b) low FL, (c) ICC1, or (d) ICC2 with medium FL conditions. • $SRMR_B$ can be used to determine model fit under following conditions: <ul style="list-style-type: none"> ○ When ICC was high (ICC3) with high FL, $SRMR_B$ was useful if NG/GS at least 100/50 or NG/GS at least 200/10. ○ When ICC was high with medium FL, $SRMR_B$ was useful (a) if NG/GS at least 100/50 or NG/GS at least 200/10 with symmetric THR; or (b) if NG/GS at least 200/10 with asymmetric THR. ○ When ICC was medium (ICC2) with high FL, $SRMR_B$ was useful if NG/GS at least 100/50 or NG/GS at least 200/10.

Note. χ_B^2 = Chi-square statistics in the between-level model. M_C = correctly specified hypothesized model. M_B = misspecification in between-level model only. NG = numbers of groups (50, 100, 200). GS = group size (10, 20, 50). ICC = intraclass correlation coefficient (ICC1 = .091, ICC2 = .231, ICC3 = .333). THR = threshold of dichotomous indicators (symmetric = 50:50 and asymmetric = 80:20). FL = factor loadings (low = .40, medium = .55, high = .70). Cut-off values for $RMSEA$, CFI , TLI , and $SRMR$ are .06, .95, .95, and .08, respectively.

Table 3 Summary of simulation findings for within-level-specific χ^2 and fit indices

Within-level-specific χ^2 and fit indices	Findings
χ^2_W	<p>M_C: GS had impacts on the values of χ^2_W but impacts of GS on the type I error rates of χ^2_W were trivial.</p> <p>M_W: NG, GS, FL, NG*FL, and GS*FL had impacts on χ^2_W. χ^2_W had stronger statistical power to correctly indicate poor models when NG, GS, and FL increased. THR was a nonignorable factor when statistical power $\geq .80$ was desired.</p> <p>Findings regarding type I error rates and statistical power:</p> <ul style="list-style-type: none"> • Type I error rates of χ^2_W were close to or below .05 across all conditions. • Statistical power of χ^2_W was close to or above .80 in following conditions: <ul style="list-style-type: none"> ○ When FL was high, NG/GS at least 50/10 can reach satisfying statistical power. ○ When FL was medium, NG/GS at least 50/10 was sufficient for symmetric THR; NG/GS at least 50/20 or at least 100/10 for asymmetric THR. ○ When FL was low, NG/GS at least 100/20 or 200/10 was sufficient for symmetric THR; NG/GS at least 100/50 or 200/20 for asymmetric THR.
$RMSEA_W$	<p>M_C: Because the variation of $RMSEA_W$ was close to zero and thus the impact of GS was self-evidently trivial.</p> <p>M_W: FL had major impacts on $RMSEA_W$, followed by THR. $RMSEA_W$ was more promising to correctly indicate poor models when FL increased or THR was symmetric.</p> <p>Findings if the cut-off value of .06 was applied:</p> <ul style="list-style-type: none"> • $RMSEA_W$ can be used to determine model fit only when FL was high and THR was symmetric.
CFI_W	<p>M_C: GS had impacts on the values of CFI_W but impacts of GS on CFI_W's sensitivity to correctly hypothesized within-level models were trivial.</p> <p>M_W: FL had impacts on the values of CFI_W but impacts of FL on CFI_W's sensitivity to poor models were trivial.</p> <p>Findings if the cut-off value of .95 was applied:</p> <ul style="list-style-type: none"> • CFI_W can be used to determine model fit in all conditions except for the conditions with low FL, NG/GS=50/10, and asymmetric THR.
TLI_W	<p>M_C: GS had impacts on the values of TLI_W but impacts of GS on TLI_W's sensitivity to correctly hypothesized within-level models were trivial.</p> <p>M_W: FL had impacts on the values of TLI_W but impacts of FL on TLI_W's sensitivity to poor models were trivial.</p> <p>Findings if the cut-off value of .95 was applied:</p> <ul style="list-style-type: none"> • TLI_W can be used to determine model fit in all conditions except for the conditions with low FL, NG/GS=50/10, and asymmetric THR.
$SRMR_W$	<p>M_C: GS had major impacts on $SRMR_W$, following by NG and THR. However, impacts of GS, NG, and THR on $SRMR_W$'s sensitivity to correctly hypothesized within-level models were trivial.</p> <p>M_W: FL had major impacts on $SRMR_W$, following by GS. However, impacts of FL and GS on $SRMR_W$'s sensitivity to poor models were manifest only when FL was low.</p> <p>Findings if the cut-off value of .08 was applied:</p> <ul style="list-style-type: none"> • $SRMR_W$ can be used to determine model fit in all conditions except for the conditions with low FL.

Note. χ^2_W = Chi-square statistics in the within-level model. M_C = correctly specified hypothesized model. M_W = misspecification in within-level model only. NG = numbers of groups (50, 100, 200). GS = group size (10, 20, 50). ICC = intraclass correlation coefficient (ICC1 = .091, ICC2 = .231, ICC3 = .333). THR = threshold of dichotomous indicators (symmetric = 50:50 and asymmetric = 80:20). FL = factor loadings (low = .40, medium = .55, high = .70). Cut-off values for $RMSEA$, CFI , TLI , and $SRMR$ are .06, .95, .95, and .08, respectively.

had a desired characteristic of indicating good model fit across replications when the between-level model was correctly specified. The η^2 s of $RMSEA_B$ were below .02, indicating that the design factors had a trivial impact on $RMSEA_B$. Figure 3 and Appendix B show that $RMSEA_B$ had means below the traditional cut-off value of .06 (Hu & Bentler, 1999) in all conditions. Alternatively, under M_B , as shown in Table 1, $RMSEA_B$ had a mean of 0.03 and an SD of 0.04. ICC ($\eta^2 = .14$) and FL ($\eta^2 = .14$) can dominantly account for the variance of $RMSEA_B$, followed by ICC*FL ($\eta^2 = .08$). According to Fig. 4 and Appendix B, we found $RMSEA_B$ was more likely to identify misspecified hypothesized between-level models when ICC or FL increased. In addition, the impact of ICC on $RMSEA_B$ was small when FL was low, and

the impact of ICC became greater when FL increased. These results show that low magnitudes of ICC or FL in models might interfere with the capacity of $RMSEA_B$ to correctly determine model fit.

We dug into tables in Appendix B to determine the extent to which $RMSEA_B$ could be used to determine model fit when the cut-off value of .06 was applied. We found $RMSEA_B$ cannot be used to determine model fit under ICC1 or low FL conditions. On the other hand, $RMSEA_B$ could be widely used in conditions with high ICC (ICC3) and high FL. In conditions with ICC3 and medium FL, increasing NG to at least 200 could prevent $RMSEA_B$ from under-rejecting poor model fit. The performance of $RMSEA_B$ is summarized in Table 2.

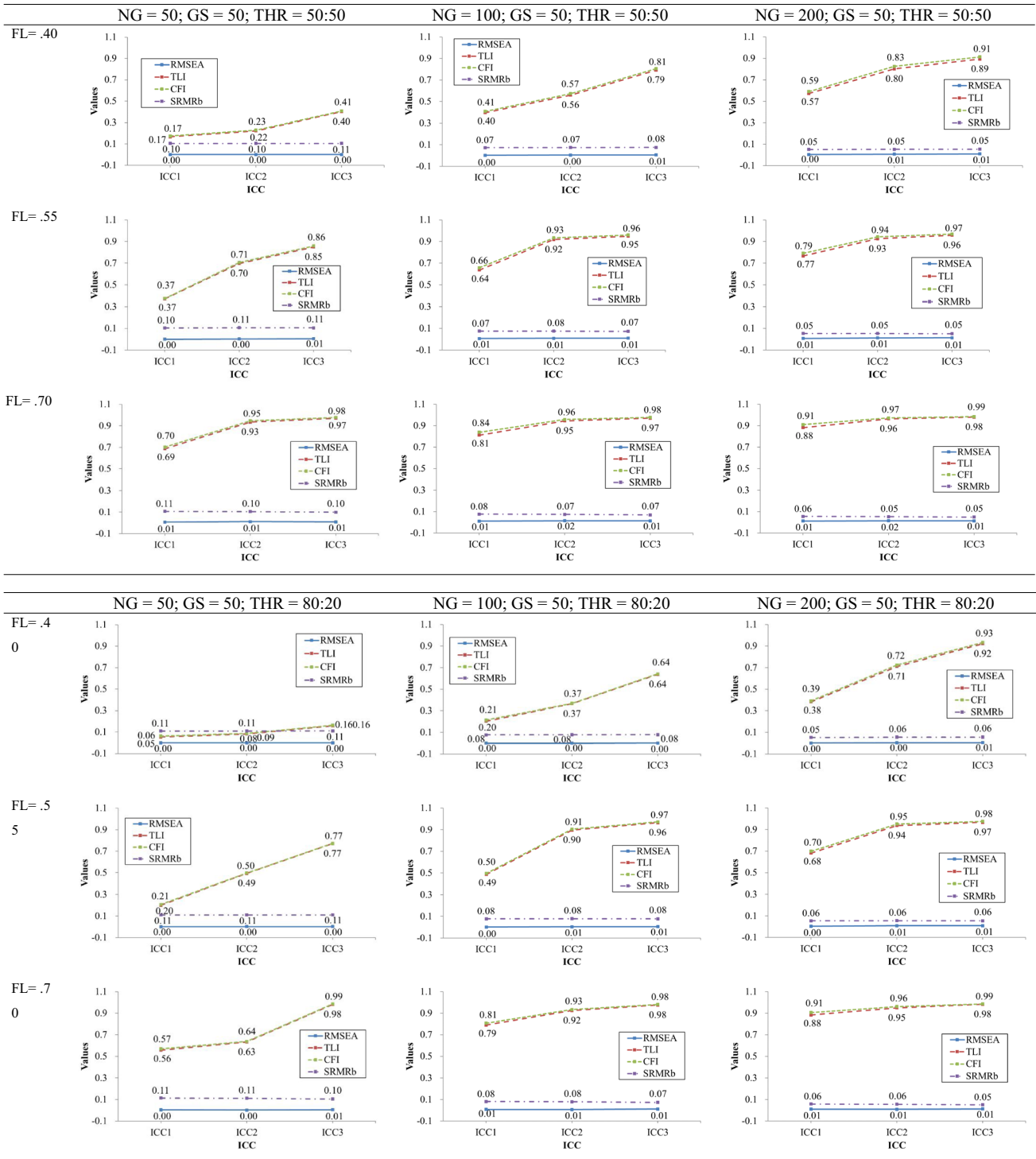


Fig. 3 Values of between-level-specific fit indices under M_C . Note. M_C = correctly specified hypothesized model. NG = numbers of groups. GS = group size. ICC = intraclass correlation coefficient. THR = threshold of dichotomous indicators (symmetric = 50:50 and asymmetric = 80:20). FL = factor loadings

Note. M_C = correctly specified hypothesized model. NG = numbers of groups. GS = group size. ICC = intraclass correlation coefficient. THR = threshold of dichotomous indicators (symmetric = 50:50 and asymmetric = 80:20). FL = factor loadings

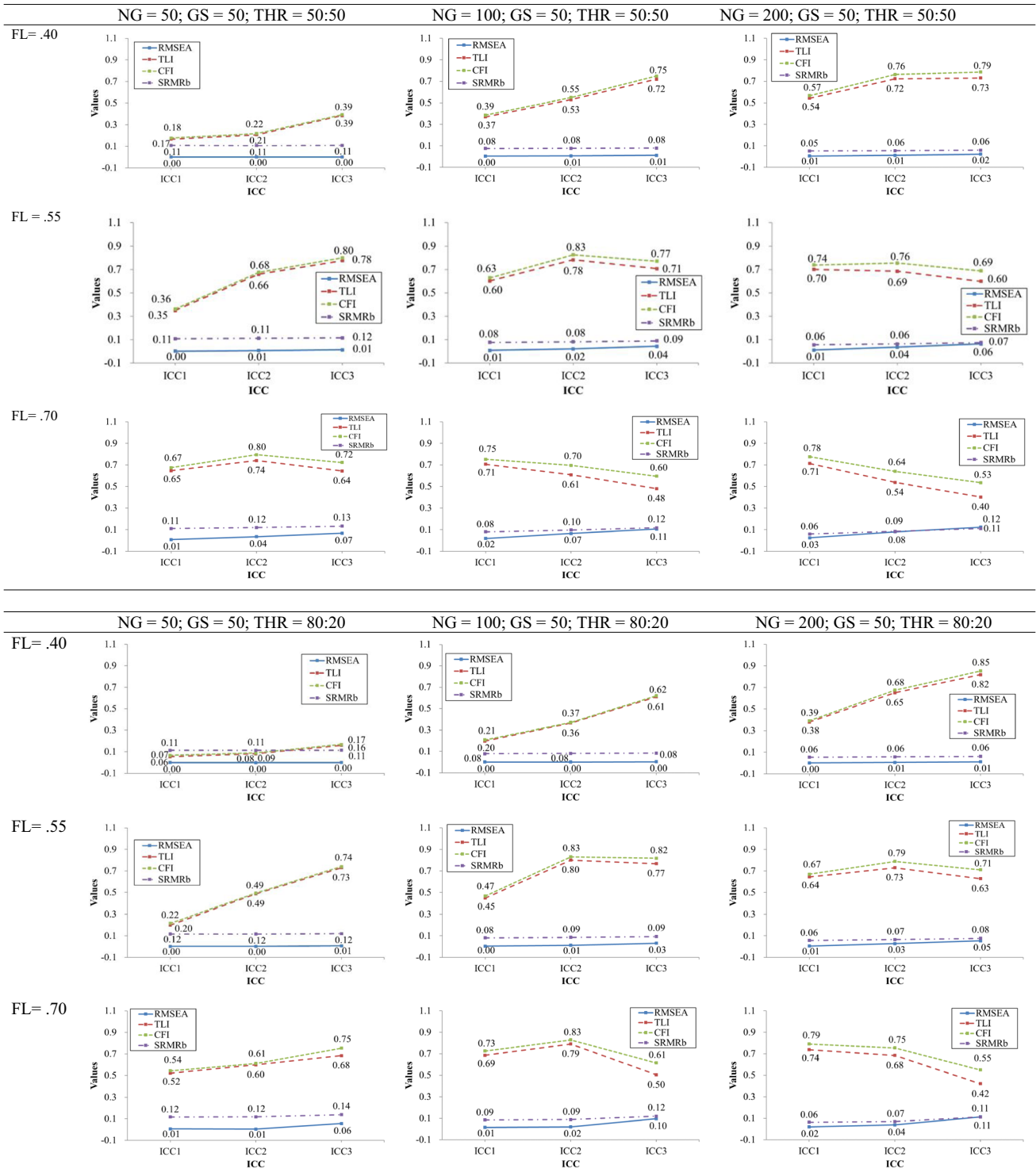


Fig. 4 Values of between-level-specific fit indices under M_B . Note. M_B = misspecification in between-level model only. NG = numbers of groups. GS = group size. ICC = intraclass correlation coefficient.

THR = threshold of dichotomous indicators (symmetric = 50:50 and asymmetric = 80:20). FL = factor loadings

CFI_B and TLI_B Under M_C , CFI_B and TLI_B were expected to be above the traditional cut-off value of .95 (Hu & Bentler, 1999). However, as reported in Table 1, CFI_B and TLI_B had

means (0.81 and 0.79, respectively) far below .95 and the SDs (0.35 and 0.36, respectively) were not negligible. This result suggested that CFI_B and TLI_B were not necessarily

promising to indicate correctly hypothesized between-level models across replications. Their η^2 's provided more information about when these two fit indices could be useful. Results show our design factors influenced CFI_B and TLI_B in a similar way. Particularly, NG ($\eta^2 = .08$ and $.07$ for CFI_B and TLI_B , respectively), ICC ($\eta^2 = .09$ for CFI_B and TLI_B), and FL ($\eta^2 = .10$ and $.09$ for CFI_B and TLI_B , respectively) had a similar magnitude of impacts on CFI_B and TLI_B . Based on Appendix B and Fig. 3, we found that both fit indices were more promising to identify correctly hypothesized between-level models when FL, ICC, or NG increased.

Under M_B , CFI_B and TLI_B also acted similarly. As presented in Table 1, CFI_B and TLI_B had means (0.65 and 0.59, respectively) far below the traditional cut-off value of .95, which were satisfying, with SD s equal to 0.34 and 0.36, respectively. The values of η^2 's ranged from .00 to .04, suggesting that design factors had minimal effects on CFI_B and TLI_B . Taken together, the results derived from M_C and M_B conditions show that low magnitudes of NG, ICC, or FL in models might limit the performance of CFI_B and TLI_B in correctly determining model fit.

We then examined tables in Appendix B to understand the performance of CFI_B and TLI_B under different conditions when the cut-off value of .95 was applied. First, we found that CFI_B and TLI_B cannot be used to determine model fit under ICC1 or low FL conditions. Second, if either ICC or FL were medium, increasing NG to at least 200 could prevent CFI_B and TLI_B from over-rejecting good model fit. Third, both indices could be widely used in conditions with ICC3, high FL, and NG at least 100. We summarized the performance of CFI_B and TLI_B in detail in Table 2.

SRMR_B Under M_C , $SRMR_B$ had a mean of 0.09, which was greater than a traditional cut-off value of .08; (Hu & Bentler, 1999), and an SD of 0.03 (see Table 1), suggesting that $SRMR_B$ might not be able to correctly indicate good model fit in all conditions. NG ($\eta^2 = .56$) accounted for a substantial proportion of $SRMR_B$'s variances followed by GS ($\eta^2 = .09$). Figure 3 and Appendix B shows that $SRMR_B$ was more promising to identify correctly hypothesized between-level models when NG or GS increased.

On the other hand, under M_B , $SRMR_B$ had a mean of 0.10 and an SD of 0.03 (see Table 1.) and it can be impacted by NG ($\eta^2 = .42$), FL ($\eta^2 = .08$), and GS ($\eta^2 = .07$). Based on Appendix B, we found $SRMR_B$ was more promising when FL increased. However, we also found $SRMR_B$ was less promising to correctly indicate poor models when NG or GS increased. After we took a close look at Figs. 3 and 4 as well as tables in Appendix B, we found this result was not surprising and could result from the fact that the patterns/values of $SRMR_B$ found under M_C were reproduced under M_B in many conditions, including low NG, low FL, ICC1, or ICC2 with medium FL [note ICC and ICC*FL Table 1

($\eta^2 = .03$, respectively) in variance of $SRMR_B$]. $SRMR_B$ performance under M_B should be interpreted cautiously. Put together, one should be aware that low magnitude of NG, FL, or ICC might cause $SRMR_B$ to lose the capacity to correctly determine model fit.

Using the tables within Appendix B, we found some conditions determining whether $SRMR_B$ could be useful when the cut-off value of .08 was applied. First, as aforementioned, we found $SRMR_B$ cannot be used to determine model fit under (a) low NG, (b) low FL, (c) ICC1, or (d) ICC2 with medium FL conditions. Second, in conditions with (a) ICC3 and high/medium FL or (b) ICC2 with high FL, $SRMR_B$ was useful if NG/GS were at least 100/50 or NG/GS were at least 200/10. When asymmetric THR appeared, NG/GS of at least 200/10 could be considered. We summarized the performance of $SRMR_B$ in Table 2.

Within-level-specific χ^2 and fit indices

χ^2_W As presented in Table 1, under M_C , χ^2_B had a mean of 31.84, an SD of 7.14, and η^2 's ranging from .00 to .09. GS was the only influential design factor ($\eta^2 = .09$); however, Appendix B shows that type I error rates of χ^2_W were close to or below .05 across all conditions. That is, GS did not significantly affect the capacity of χ^2_W to identify correctly hypothesized within-level models. In contrast, under M_W , χ^2_W had a mean of 306.75 and an SD of 377.37, as shown in Table 1. Large variations in values of χ^2_W resulted from FL ($\eta^2 = .14$), NG ($\eta^2 = .11$), GS ($\eta^2 = .08$), NG*FL ($\eta^2 = .06$), and GS*FL ($\eta^2 = .06$). Based on Appendix B, we found χ^2_W had higher statistical power (close to or above .80) to detect misspecified within-level models when NG, GS, and FL increased. The above results show when using χ^2_W , one should be aware that type I error rates might not be a problem, but low statistical power might appear when NG, GS, or FL were low.

We examined Appendix B to uncover the conditions where χ^2_W had statistical power close to or above .80. First, when FL was high, χ^2_W was able to reach satisfying statistical power given our minimum combination of NG/GS (50/10). Second, when FL was medium, thresholds of dichotomous indicators started to slightly weigh into the requirement of NG and GS (THR had a η^2 of .03, as shown in Table 1): NG/GS = 50/10 was sufficient for symmetric THR, while NG/GS of at least 50/20 or at least 100/10 was required for asymmetric THR. Third, when FL was low, THR could weigh in a little more: NG/GS of at least 100/20 or 200/10 was sufficient for symmetric THR; NG/GS of at least 100/50 or 200/20 for asymmetric THR. The performance of χ^2_W is summarized in Table 3.

RMSEA_W Table 1 shows that under M_C , $RMSEA_W$ had a mean equal to 0.00 and an SD equal to 0.01, suggesting that

$RMSEA_W$ shows a desired property of indicating good model fit across replications when the within-level model was correctly specified. GS ($\eta^2 = .10$) was the only factor that significantly influenced $RMSEA_W$. Nevertheless, the impact of GS on $RMSEA_W$ could be ignored because the variation of $RMSEA_W$ was close to zero; thus the impact of GS was self-evidently trivial. $RMSEA_W$ had means below a traditional cut-off value of .06 (Hu & Bentler, 1999) in all conditions.

On the other hand, under M_B , as shown in Table 1, $RMSEA_W$ had a mean of 0.04, an SD of 0.02. FL ($\eta^2 = .56$) dominantly affected $RMSEA_W$, followed by THR ($\eta^2 = .09$). The tables in Appendix B show that $RMSEA_W$ was more likely to identify misspecified hypothesized within-level models when FL was increased or THR was symmetric. After examining tables in Appendix B, we found $RMSEA_W$ tended to under-reject poor model fit when the cut-off value of .06 was applied. Particularly, results show $RMSEA_W$ can be used to determine model fit in very restricted conditions: when FL was high and THR was symmetric. The performance of $RMSEA_W$ is summarized in Table 3.

CFI_W and TLI_W In general, CFI_W and TLI_W performed similarly. As presented in Table 1, under M_C , both CFI_W and TLI_W had a mean equal to 0.99 and an SD equal to 0.03, suggesting that these two fit indices were able to correctly indicate good model fit across replications. The η^2 of GS was .07 for CFI_W and .08 for TLI_W ; however, the impact of GS could be neglected due to a small variation in CFI_W and TLI_W . Under M_w , CFI_W and TLI_W had means of 0.74 and 0.66, and SD s of 0.08 and 0.10, respectively. Although FL seemed to affect CFI_W and TLI_W ($\eta^2 = .09$), Appendix B reveals that both CFI_W and TLI_W had means far below a traditional cut-off value of .95 (Hu & Bentler, 1999) in most conditions, suggesting that these two fit indices were promising to identify incorrectly hypothesized within-level models.

We examined Appendix B to understand the sensitivity of CFI_W and TLI_W under different conditions when the cut-off value of .95 was applied. We found that CFI_W and TLI_W can be widely used with the cut-off value of .95 except for the conditions with low FL, NG/GS = 50/10, and asymmetric THR. We summarized the performance of CFI_W and TLI_W in Table 3.

$SRMR_W$ Under M_C , $SRMR_W$ had a mean of 0.04, which was below a traditional cut-off value of .08 (Hu & Bentler, 1999), and a small SD of 0.02 (shown in Table 1). The results suggest that $SRMR_W$ showed promise to correctly indicate good model fit in most replications. NG ($\eta^2 = .29$), GS ($\eta^2 = .40$), and THR ($\eta^2 = .08$), were practically significant but could be ignored due to a small variation in $SRMR_W$. On the other hand, under M_B , $SRMR_W$ had a mean of 0.09, which was above a traditional cut-off value of .08, and a small SD of 0.02, suggesting that $SRMR_W$ was useful to indicate poor

model fit in most replications. The influence of FL ($\eta^2 = .54$) and GS ($\eta^2 = .06$) could be ignored since the variation in $SRMR_W$ was trivial. Using Appendix B, we found $SRMR_W$ with a cut-off value of .08 can be used to determine model fit in all conditions except for the conditions with low FL. We summarized the performance of $SRMR_W$ in Table 3.

Discussion

The current study examined the performance of level-specific χ^2 test statistics and fit indices derived from the partially saturated-model method in correctly identifying good or poor model fit in MCFA with dichotomous indicators. Tables 2 and 3 summarize our simulation results that could inform applied researchers of their usage of level-specific χ^2 test statistics and fit indices in evaluating MCFA with dichotomous indicators. Note that the results were tied to the specific conditions of this research and should not be overgeneralized. In this section, we highlight major findings for discussion.

First, as summarized in Table 2, the performance of b-l-s χ^2 and fit indices was highly associated with the magnitude of ICC in data. Both b-l-s χ^2 and fit indices were more promising indicators to correctly indicate model fit in MCFA with dichotomous indicators when ICC increased. Unfortunately, in conditions with low ICC (ICC1 = .091), we found χ^2_B had statistical power far below .80, and all b-l-s fit indices ($RMSEA_B$, CFI_B , TLI_B , and $SRMR_B$) were not useful to evaluate between-level models along with traditional cut-off values. This is congruent with previous findings derived from MCFA with continuous indicators (Hsu et al., 2017; Lee & Sohn, 2022). Note that between-level indicators have stronger relations when ICC is higher, allowing for possibly “greater” discrepancies between the model-implied and observed between-level variance-covariance matrices (Kline, 2011). As a result, given identical misspecification and sample size, the value of χ^2_B increased (higher statistical power to detect the misspecification) with higher ICC. Because $RMSEA_B$ is a function of the χ^2_B , the pattern of $RMSEA_B$ in response to ICC was similar to χ^2_B . Regarding CFI_B and TLI_B , both were functions of the distance between the χ^2 value of the hypothesized model and that of the baseline null model—the greater the distance between two χ^2 values, the larger were CFI_B and TLI_B . When the between-level model was misspecified, a higher ICC not only led to greater χ^2 values of the hypothesized model and the baseline null model, but also resulted in larger distances between these two χ^2 values. That is why we observed that both CFI_B and TLI_B were more promising for detecting misspecified between-level models when ICC increased. Similarly, $SRMR_B$ was more effective when ICC was higher. This is because $SRMR_B$ reflects the deviation between the

between-level model-implied and observed variance-covariance matrices; the deviation would become larger with an increase in ICC. On the other hand, we found that ICC did not affect the effectiveness of w-l-s χ^2 and fit indices because the within-level variation was held constant across the ICC conditions in this study (i.e., within-level indicators had constant relations across ICC conditions).

Second, as presented in Table 2, another factor influencing the performance of b-l-s χ^2 and fit indices was the magnitude of factor loadings. Decreasing factor loadings were associated with decreasing value of χ_B^2 and of all b-l-s fit indices. In conditions with low FL (.40), the statistical power of χ_B^2 was far below .80, and no b-l-s fit indices can correctly identify model fit along with traditional cut-off values. These findings are in line with Nestler's (2013) and Heene et al.'s (2011) studies. As noted by Heene et al. (2011), the problem of using χ_B^2 , $RMSEA_B$, and $SRMR_B$ in low FL conditions is that the values of these indicators will be too small to detect the misspecified between-level models. On the other hand, the problem of using CFI_B and TLI_B in low FL conditions is that these indicators will over-reject correctly hypothesized between-level models.

Third, in addition to ICC and FL, the size of NG played an important role in the performance of χ_B^2 as well as b-l-s fit indices if traditional cut-off values were applied. As shown in Table 2, a small to medium NG (50–100) might be sufficient for b-l-s χ^2 and fit indices only if both ICC and FL were high. In the remaining conditions, an NG of 200 was needed to use b-l-s χ^2 and fit indices for between-level model evaluation. These findings raise a pressing issue—when both ICC and FL are not high, how can we assess the between-level model fit in MCFA with dichotomous indicators given an NG far less than 200? Thus far, the possible solution we are aware of is to compute adjusted χ_B^2 test statistics based on the post hoc corrections (Herzog & Boomsma, 2009; Savalei, 2010). Specifically, when the NG is small (i.e., small sample size at the between-level), the estimated χ_B^2 might not follow χ^2 distribution (Bentler & Yuan, 1999). Previous research has shown that computing adjusted χ^2 is promising in providing sufficient statistical power and reliable model fit for data with nonnormality or small sample size (Savalei, 2010). To the best of our knowledge, the effectiveness of adjusted χ_B^2 test statistics based on the post hoc corrections in small NG scenarios has not been studied in MCFA; future studies are needed to shed new light on this issue. In addition, little is known about whether b-l-s fit indices computed based on adjusted χ_B^2 are useful for detecting misspecified between-level models in MCFA. Future studies are encouraged to explore this topic.

Fourth, as shown in Table 1, THR had close-to-zero η^2 s for b-l-s χ^2 and fit indices, suggesting the impact of asymmetric thresholds (80:20) on the performance of b-l-s χ^2 and fit indices might be trivial. Still, we did discover that the impact of THR can slightly weigh into the performance of

b-l-s χ^2 and fit indices when ICC was not high. For example, as summarized in Table 2, when ICC was medium (ICC2), we can reach a satisfying statistical power of χ_B^2 only in conditions with high FL, large NG (200), and symmetric THR, but not in any conditions with asymmetric THR. This finding is in line with Rhemtulla et al. (2012) and Savalei and Rhemtulla (2013), who found that asymmetric thresholds and small samples reduced adjusted χ^2 statistical power. In addition, we observed that under ICC2, b-l-s fit indices might need larger NG/GS to compensate for the presence of asymmetric thresholds. For instance, as shown in Table 2, in conditions with ICC2 and high FL, CFI_B and TLI_B required NG/GS of at least 100/50 or NG/GS of at least 200/10 if thresholds were symmetric, but CFI_B and TLI_B would require NG/GS at least 200/50 if thresholds were asymmetric. This finding might inform applied researchers of the sample size requirement when asymmetric thresholds occur.

Fifth, the magnitude of factor loadings and asymmetric thresholds should be considered when using w-l-s χ^2 and fit indices. As presented in Table 3, w-l-s χ^2 and fit indices except for $RMSEA_B$ were useful in most conditions to determine model fit at within-level in MCFA with dichotomous indicators. Our study complements earlier work (Hsu, 2009; Hsu et al., 2017; Ryu & West, 2009) that found that w-l-s χ^2 and fit indices performed well in MCFA with continuous indicators, and our findings further support the adequacy of adopting w-l-s χ^2 and fit indices to evaluate within-level models in MCFA if indicators are dichotomous. However, we found that both FL and THR could influence the performance of w-l-s χ^2 and fit indices. As aforementioned, decreasing factor loadings were associated with decreasing value of χ^2 and of fit indices. As a result, one would need to increase NG/GS to compensate for the impact of low factor loadings. In addition, we found THR had a slight impact and could weigh in the performance of χ_W^2 , $RMSEA_W$, CFI_W , and TLI_W . Table 3 suggested $RMSEA_W$ was heavily affected by FL and THR: $RMSEA_W$ can be used to determine model fit only when FL was high and THR was symmetric. The good news is that others fit indices were useful for determining within-level model fit along with traditional cut-off values.

Last but not least, the results suggest that a low ICC (.091) was linked to the low convergence rates in MCFA with dichotomous indicators. This finding was in line with Hox and Maas's (2001) and Hsu et al.'s (2017) study findings that were based on an MCFA population model with continuous indicators. Due to the fact that ICCs around .10 seem to be very common in applied research. Explore alternative estimation strategies (e.g., Bayesian estimation; see Depaoli & Clifton, 2015) to resolve the issue (i.e., low convergence rates) for multilevel confirmatory factor analysis can be crucial. In other words, stable and reliable estimation methods need to be developed and evaluated before addressing further questions about model fit.

Several limitations should be noted. First, our findings should be generalized only to studies that apply MCFA models with dichotomous indicators (Fig. 1). Further studies are needed to test the replicability of our results using different models (e.g., structural models). Second, only a limited number of design factors were considered in the current study. Future studies could expand the examination of additional design factors, such as unequal factor loadings, unequal group conditions, the number of observed indicators per latent factor, and different types of misspecifications (e.g., misspecified in cross-loadings; Moshagen & Auerswald, 2018). Third, considering the computational complexity in MCFA with dichotomous indicators, we generated 500 replications for each of 162 conditions and investigated one misspecification condition in this study which restricted the ability to generalize. Future research could extend our investigation using a more powerful computer.

Concluding remarks

According to our simulation results, we recommend that practitioners should be aware that the performance of b-l-s χ^2 and fit indices was mainly influenced by data ICC and factor loadings (FL), followed by number of groups (NG), while thresholds of dichotomous indicators (THR) could slightly weigh in the performance of b-l-s fit indices in some conditions. Specifically, both b-l-s χ^2 and fit indices were more promising indicators to correctly indicate model fit in MCFA with dichotomous indicators when ICC or FL increased. In conditions with low ICC (.091) or low FL (.40), χ^2_B had statistical power far below .80, and all b-l-s fit indices were not useful to evaluate between-level models along with traditional cut-off values. In terms of NG, we found a small to medium NG (50–100) might be sufficient for b-l-s χ^2 and fit indices only if both ICC and factor loadings were high, while in remaining conditions, an NG of 200 was needed to be able to use b-l-s χ^2 and fit indices for between-level model evaluation. The performance of b-l-s χ^2 and fit indices is summarized in Table 2. On the other hand, we recommend that practitioners use w-l-s χ^2 and fit indices (except for $RMSEA_w$) along with traditional cut-off values to evaluate within-level models comprising dichotomous indicators. However, practitioners should be aware that both FL and THR could influence the performance of w-l-s χ^2 and fit indices. Both w-l-s χ^2 and of fit indices were more promising to determine model fit when FL increased. When FL was low, one would need to increase NG and group size (GS) to compensate for the impact of low factor loadings. THR had a slight impact and could weigh in the performance of χ^2_w , $RMSEA_w$, CFI_w , and TLI_w . Unfortunately, $RMSEA_w$ was heavily affected by FL and THR: $RMSEA_w$ can be used to determine model fit only when FL was high and THR was

symmetric. The performance of w-l-s χ^2 and fit indices is summarized in Table 3.

Appendix A. The formulas for the level-specific fit indexes

Chi-square statistic

By applying the PS method, the $\chi^2_{PS_B}$ can be defined as follows:

$$\chi^2_{PS_B} = F_{ML}[\Sigma_B(\hat{\theta}), \Sigma_W(\hat{\theta}_S)] - F_{ML}[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta}_S)] \quad (A1)$$

in which the $F_{ML}[\Sigma_B(\hat{\theta}), \Sigma_W(\hat{\theta}_S)]$ is the fitting function for the saturated within-level model, while the $F_{ML}[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta}_S)]$ is the fitting function when both within-level and between-level levels are saturated, i.e., the fully saturated model. The degrees of freedom of $\chi^2_{PS_B}$, denoted by df_{PS_B} , equals to the difference between the number of parameters in the partially saturated model and the fully saturated model and can be defined in terms of A2 as follows:

$$df_{PS_B} = df_{B, Saturated} - df_{B, Hypothesized} \quad (A2)$$

in which $df_{B, Saturated}$ and $df_{B, Hypothesized}$ represent number of parameters in the saturated within-level model and fully saturated model, respectively.

Similarly, the $\chi^2_{PS_W}$ and its corresponding degree of freedom df_{PS_W} can be derived in terms of A3 and A4, respectively, as follows:

$$\chi^2_{PS_W} = F_{ML}[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta})] - F_{ML}[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta}_S)] \quad (A3)$$

$$df_{PS_W} = df_{W, Saturated} - df_{W, Hypothesized} \quad (A4)$$

RMSEA

Given $\chi^2_{PS_B}$ and df_{PS_B} , the $RMSEA$ for the between-level model, $RMSEA_{PS_B}$, can be derived in terms of A5 as follows:

$$RMSEA_{PS_B} = \sqrt{\text{Max} \left(\frac{\chi^2_{PS_B} - df_{PS_B}}{df_{PS_B}(J)}, 0 \right)} \quad (A5)$$

The J in A5 is the sample size (number of groups) for the between-level model. To reduce the random error component of F_{ML} , the non-central parameter df_{PS_B} is subtracted from the $\chi^2_{PS_B}$ (Rigdon, 1996). Therefore, $\chi^2_{PS_B} - df_{PS_B}$ stands for an unbiased estimator of F_{ML} . The denominator J in A5 functions as a penalty for lack of parsimony. The df_{PS_B} in the denominator transfers $(\chi^2_{PS_B} - df_{PS_B})/J$ into a measure of lack of fit per df . In addition, the $RMSEA_{PS_B}$

is is set to zero providing $\chi^2_{PS_B}$ is smaller than df_{PS_B} . Similarly, the $RMSEA_{PS_W}$ can be derived from A6, in which N denotes the total sample size for the within-level model:

$$RMSEA_{PS_W} = \sqrt{\text{Max} \left(\frac{\chi^2_{PS_W} - df_{PS_W}}{df_{PS_W}(N - J)}, 0 \right)} \tag{A6}$$

CFI

CFI is one incremental fit index that can be used to evaluate the goodness-of-fit by comparing the hypothesized model with the independence model (Bentler, 1990). The CFI for the between-level model (CFI_{PS_B}), derived from those proposed by Ryu and West (2009), is showed as follows:

$$CFI_{PS_B} = 1 - \frac{\text{Max} \left[\left(\chi^2_{PS_B} - df_{PS_B} \right), 0 \right]}{\text{Max} \left[\left(\chi^2_{I_{B,S,W}} - df_{I_{B,S,W}} \right), 0 \right]} \tag{A7}$$

in which $\chi^2_{I_{B,S,W}}$ represents the χ^2 test statistic with an independence (null) between-level model and a saturated within-level model. The $\chi^2_{I_{B,S,W}}$ can be derived in term of A8 as follows:

$$\chi^2_{I_{B,S,W}} = F_{ML} \left[\Sigma_B(\hat{\theta}_1), \Sigma_W(\hat{\theta}_S) \right] - F_{ML} \left[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta}_S) \right] \tag{A8}$$

The independence model aforementioned can be derived by correlating all the observed variables and constraining all the correlation coefficients to zero. The corresponding $df_{I_{B,S,W}}$ is defined as:

$$df_{I_{B,S,W}} = df_{B,Saturated} - df_{B,Independent} \tag{A9}$$

The CFI_{PS_B} measures the goodness-of-fit by comparing the hypothesized model with the independence model at the between level, when the within-level model is saturated. Similarly, the CFI_{PS_W} can be derived in terms of A10 as follows:

$$CFI_{PS_W} = 1 - \frac{\text{Max} \left[\left(\chi^2_{PS_W} - df_{PS_W} \right), 0 \right]}{\text{Max} \left[\left(\chi^2_{S_{B,I,W}} - df_{S_{B,I,W}} \right), 0 \right]} \tag{A10}$$

where $\chi^2_{S_{B,I,W}}$ showed in A11 represents the χ^2 statistics with a saturated between-level model and an independence within-level model. The corresponding $df_{S_{B,I,W}}$ can be derived in terms of A12 as follows:

$$\chi^2_{S_{B,I,W}} = F_{ML} \left[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta}_I) \right] - F_{ML} \left[\Sigma_B(\hat{\theta}_S), \Sigma_W(\hat{\theta}_S) \right] \tag{A11}$$

$$df_{S_{B,I,W}} = df_{W,Saturated} - df_{W,Independent} \tag{A12}$$

TLI

The TLI is classified as a non-normed fit index that penalizes for lack of parsimony of the model (Tucker &

Lewis, 1973). The TLI_{PS_B} evaluates the goodness-of-fit of the between-level model by comparing the hypothesized between-level model and the independent between-level model when the within-level model is saturated. In contrast, the TLI_{PS_W} estimates the goodness-of-fit of the within-level model by comparing the hypothesized within-level model and the independent within-level model when the between-level model is saturated. The formulas for TLI_{PS_B} and TLI_{PS_W} could be derived in terms of A13 and A14 respectively, as follows:

$$TLI_{PS_B} = \frac{\frac{\chi^2_{I_{B,S,W}}}{df_{I_{B,S,W}}} - \frac{\chi^2_{PS_B}}{df_{PS_B}}}{\frac{\chi^2_{I_{B,S,W}}}{df_{I_{B,S,W}}} - 1} \tag{A13}$$

$$TLI_{PS_W} = \frac{\frac{\chi^2_{S_{B,I,W}}}{df_{S_{B,I,W}}} - \frac{\chi^2_{PS_W}}{df_{PS_W}}}{\frac{\chi^2_{S_{B,I,W}}}{df_{S_{B,I,W}}} - 1} \tag{A14}$$

SRMR

SRMR can be classified into the $SRMR_W$ at the within-level and the $SRMR_B$ at the between-level models, respectively. Note that SRMR is independent of χ^2 statistics and can be defined as the deviation between the sample variance-covariance matrix and the reproduced variance-covariance matrix. Statistical packages, such as Mplus (Muthen & Muthen, 2017), can provide $SRMR_W$ and $SRMR_B$ for model evaluations. Moreover, $SRMR_B$ can be regarded as the normed average distance between the sample variance matrix at the between level. Therefore, the $SRMR_B$ can be derived in terms of A15 as follows:

$$SRMR_B = \sqrt{\frac{2 \sum_{i=1}^p \sum_{j=1}^i \left[\frac{(\Sigma_{B_{ij}} - \Sigma_B(\theta)_{ij})^2}{\Sigma_{B_{ii}} \Sigma_{B_{jj}}} \right]}{p(p+1)}} \tag{A15}$$

Using the similar definition, $SRMR_W$ can be defined as the normed average distance between the sample variance matrix and model-implied variance matrix at the within level. Therefore, the $SRMR_W$ can be derived in terms of A16 as follows:

$$SRMR_W = \sqrt{\frac{2 \sum_{i=1}^p \sum_{j=1}^i \left[\frac{(\Sigma_{W_{ij}} - \Sigma_W(\theta)_{ij})^2}{\Sigma_{W_{ii}} \Sigma_{W_{jj}}} \right]}{p(p+1)}} \tag{A16}$$

Appendix B. Results of the additional simulations

Table 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21.

Table 4 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC1 (.091) for THR = 50:50 and FL=.40

	NG =								
	50		100		200				
	10	20	50	10	50	10	20	50	
M_C									
<i>B-I-s</i>									
χ^2_B (<i>df</i> = 34)	28.80 (4.77)	29.62 (4.72)	25.62 (3.45)	28.20 (5.11)	27.56 (4.71)	27.65 (5.92)	30.75 (6.74)	29.44 (6.18)	
Type I error rate	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02	
<i>RMSEA_B</i>	0.00 (0.01)	0.01 (0.02)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)	
<i>CFI_B</i>	0.38 (0.47)	0.42 (0.46)	0.17 (0.37)	0.39 (0.47)	0.41 (0.47)	0.48 (0.48)	0.67 (0.42)	0.59 (0.46)	
<i>TLL_B</i>	0.37 (0.47)	0.40 (0.46)	0.17 (0.37)	0.38 (0.46)	0.40 (0.47)	0.47 (0.48)	0.64 (0.43)	0.57 (0.46)	
<i>SRMR_B</i>	0.13 (0.02)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)	0.05 (0.01)	
<i>W-I-s</i>									
χ^2_W (<i>df</i> = 34)	33.87 (6.01)	31.61 (5.23)	24.46 (3.86)	33.57 (6.25)	26.25 (4.86)	33.00 (7.53)	33.33 (7.28)	29.75 (6.53)	
Type I error rate	0.02	0.00	0.00	0.00	0.00	0.05	0.02	0.01	
<i>RMSEA_W</i>	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	
<i>CFI_W</i>	0.95 (0.09)	0.98 (0.04)	1.00 (0.00)	0.97 (0.03)	1.00 (0.00)	0.99 (0.03)	0.99 (0.01)	1.00 (0.00)	
<i>TLL_W</i>	0.93 (0.12)	0.98 (0.05)	1.00 (0.00)	0.97 (0.04)	1.00 (0.00)	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	
<i>SRMR_W</i>	0.07 (0.01)	0.05 (0.01)	0.03 (0.00)	0.05 (0.01)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)	
M_B									
<i>B-I-s</i>									
χ^2_B (<i>df</i> = 35)	30.84 (5.06)	31.50 (4.84)	27.32 (3.69)	30.16 (5.29)	29.43 (4.97)	29.82 (6.29)	32.79 (6.87)	31.35 (6.45)	
Power	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.02	
<i>RMSEA_B</i>	0.01 (0.02)	0.01 (0.02)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	
<i>CFI_B</i>	0.34 (0.44)	0.39 (0.44)	0.18 (0.37)	0.36 (0.44)	0.39 (0.46)	0.45 (0.45)	0.64 (0.41)	0.57 (0.45)	
<i>TLL_B</i>	0.31 (0.44)	0.36 (0.44)	0.17 (0.37)	0.34 (0.44)	0.37 (0.45)	0.42 (0.46)	0.61 (0.42)	0.54 (0.45)	
<i>SRMR_B</i>	0.13 (0.02)	0.11 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.06 (0.01)	0.06 (0.01)	0.05 (0.01)	
<i>W-I-s</i>									
χ^2_W (<i>df</i> = 35)	44.84 (9.43)	49.63 (9.94)	50.05 (9.38)	57.76 (12.71)	102.65 (18.27)	82.78 (16.61)	135.77 (23.65)	236.11 (30.84)	
Power	0.27	0.45	0.49	0.69	1.00	1.00	1.00	1.00	
<i>RMSEA_W</i>	0.02 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.00)	0.03 (0.00)	0.03 (0.00)	0.02 (0.00)	
<i>CFI_W</i>	0.77 (0.18)	0.79 (0.11)	0.84 (0.08)	0.75 (0.12)	0.76 (0.06)	0.74 (0.08)	0.73 (0.05)	0.74 (0.04)	
<i>TLL_W</i>	0.70 (0.23)	0.73 (0.15)	0.80 (0.10)	0.68 (0.16)	0.70 (0.07)	0.67 (0.10)	0.66 (0.07)	0.67 (0.05)	
<i>SRMR_W</i>	0.08 (0.01)	0.06 (0.01)	0.05 (0.01)	0.07 (0.01)	0.05 (0.00)	0.06 (0.01)	0.05 (0.00)	0.05 (0.00)	

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. *B-I-s* = Between-level specific. *W-I-s* = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 5 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC1 (.091) for THR = 50:50 and FL = .55

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 34)	28.91 (4.74)	30.18 (4.57)	27.28 (3.84)	29.43 (5.35)	30.35 (6.02)	29.98 (5.30)	28.72 (6.08)	31.05 (6.55)	31.69 (6.50)						
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03						
$RMSEA_B$	0.01 (0.01)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)						
CFI_B	0.41 (0.47)	0.56 (0.46)	0.37 (0.48)	0.55 (0.46)	0.61 (0.44)	0.66 (0.43)	0.60 (0.46)	0.72 (0.39)	0.79 (0.34)						
TLL_B	0.39 (0.47)	0.54 (0.46)	0.37 (0.48)	0.53 (0.47)	0.58 (0.45)	0.64 (0.44)	0.58 (0.46)	0.69 (0.40)	0.77 (0.35)						
$SRMR_B$	0.13 (0.02)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)	0.05 (0.01)						
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 34)	33.30 (6.15)	32.49 (5.97)	26.76 (4.55)	33.32 (6.20)	33.92 (7.23)	28.62 (5.46)	33.44 (6.98)	34.82 (7.67)	31.18 (7.41)						
Type I error rate	0.01	0.01	0.00	0.02	0.03	0.00	0.02	0.05	0.03						
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)						
TLL_W	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)						
$SRMR_W$	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)						
M_B															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 35)	31.12 (4.92)	32.38 (5.02)	29.21 (3.97)	31.49 (5.39)	32.67 (6.11)	32.15 (5.37)	31.42 (6.28)	33.68 (6.96)	34.50 (7.15)						
Power	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)						
CFI_B	0.38 (0.44)	0.52 (0.45)	0.36 (0.47)	0.51 (0.45)	0.55 (0.42)	0.63 (0.43)	0.56 (0.45)	0.67 (0.38)	0.74 (0.34)						
TLL_B	0.35 (0.44)	0.49 (0.45)	0.35 (0.47)	0.49 (0.46)	0.51 (0.44)	0.60 (0.44)	0.54 (0.45)	0.64 (0.40)	0.70 (0.36)						
$SRMR_B$	0.13 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)						
M_W															
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 35)	68.71 (15.40)	98.08 (19.72)	135.43 (26.49)	118.04 (26.17)	203.04 (32.77)	353.75 (46.51)	209.67 (33.79)	399.39 (47.57)	843.26 (71.33)						
Power	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.04 (0.01)	0.04 (0.01)	0.03 (0.00)	0.05 (0.01)	0.05 (0.00)	0.04 (0.00)	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)						
CFI_W	0.73 (0.11)	0.74 (0.07)	0.75 (0.05)	0.71 (0.08)	0.71 (0.05)	0.73 (0.03)	0.71 (0.05)	0.72 (0.03)	0.72 (0.02)						
TLL_W	0.66 (0.14)	0.66 (0.09)	0.68 (0.06)	0.63 (0.11)	0.63 (0.06)	0.65 (0.04)	0.63 (0.07)	0.64 (0.04)	0.64 (0.03)						
$SRMR_W$	0.10 (0.01)	0.08 (0.01)	0.08 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.00)	0.08 (0.01)	0.08 (0.00)	0.08 (0.00)						

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 6 Means of level-specific χ^2 test statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC1 (.091) for THR = 50:50 and FL = .70

	100						200					
	50		20		50		10		20		50	
	10	50	10	50	10	50	10	50	10	20	50	
M_C												
<i>B-l-s</i>												
χ^2_B (<i>df</i> = 34)	30.15 (4.33)	31.02 (5.05)	30.16 (4.61)	30.25 (5.61)	31.89 (6.34)	32.31 (5.84)	31.13 (6.88)	33.07 (7.22)	34.28 (7.04)			
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.02	0.03			
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)			
CFI_B	0.54 (0.47)	0.64 (0.44)	0.70 (0.43)	0.69 (0.42)	0.76 (0.35)	0.84 (0.29)	0.81 (0.32)	0.88 (0.22)	0.91 (0.14)			
TLL_B	0.52 (0.48)	0.61 (0.44)	0.69 (0.44)	0.67 (0.43)	0.73 (0.37)	0.81 (0.31)	0.78 (0.35)	0.85 (0.24)	0.88 (0.18)			
$SRMR_B$	0.15 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)			
<i>W-l-s</i>												
χ^2_W (<i>df</i> = 34)	34.14 (6.55)	34.75 (6.63)	29.30 (5.16)	34.18 (6.40)	34.12 (7.36)	31.69 (6.67)	34.03 (7.34)	33.83 (7.60)	32.83 (7.56)			
Type I error rate	0.02	0.04	0.00	0.02	0.04	0.01	0.03	0.04	0.04			
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)			
CFI_W	1.00 (0.02)	1.00 (0.09)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)			
TLL_W	1.00 (0.02)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)			
$SRMR_W$	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)			
M_B												
<i>B-l-s</i>												
χ^2_B (<i>df</i> = 35)	32.35 (4.47)	33.39 (5.18)	32.27 (4.63)	32.72 (5.88)	35.11 (6.89)	35.94 (6.32)	34.73 (7.20)	38.16 (8.50)	41.02 (8.62)			
Power	0.00	0.00	0.00	0.00	0.02	0.03	0.02	0.09	0.16			
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.02 (0.02)	0.03 (0.03)			
CFI_B	0.50 (0.45)	0.60 (0.43)	0.67 (0.42)	0.64 (0.42)	0.68 (0.36)	0.75 (0.30)	0.74 (0.33)	0.78 (0.25)	0.78 (0.20)			
TLL_B	0.46 (0.46)	0.56 (0.44)	0.65 (0.44)	0.61 (0.44)	0.63 (0.38)	0.71 (0.33)	0.70 (0.36)	0.73 (0.30)	0.71 (0.25)			
$SRMR_B$	0.15 (0.02)	0.13 (0.02)	0.11 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)			
M_W												
<i>W-l-s</i>												
χ^2_W (<i>df</i> = 35)	118.98 (26.38)	206.451 (36.51)	383.47 (55.46)	220.81 (42.71)	426.62 (57.20)	916.11 (87.51)	431.82 (62.09)	881.51 (81.86)	2071.27 (135.84)			
Power	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
$RMSEA_W$	0.07 (0.01)	0.07 (0.01)	0.06 (0.06)	0.07 (0.01)	0.08 (0.01)	0.07 (0.00)	0.08 (0.01)	0.08 (0.00)	0.08 (0.00)			
CFI_W	0.70 (0.08)	0.70 (0.06)	0.72 (0.04)	0.70 (0.07)	0.70 (0.04)	0.71 (0.03)	0.69 (0.05)	0.70 (0.03)	0.70 (0.02)			
TLL_W	0.62 (0.11)	0.62 (0.08)	0.64 (0.05)	0.61 (0.09)	0.61 (0.05)	0.63 (0.03)	0.61 (0.06)	0.61 (0.04)	0.62 (0.02)			
$SRMR_W$	0.12 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)			

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-l-s = Between-level specific. W-l-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 test statistics are reported to two decimal places for the sake of space.

Table 7 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC1 (.091) for THR = 80:20 and FL = .40

	100						200							
	50		20		50		10		50		10		50	
	NG =	GS =	NG =	GS =	NG =	GS =	NG =	GS =	NG =	GS =	NG =	GS =	NG =	GS =
M_C														
<i>B-I-s</i>														
χ^2_B (<i>df</i> = 34)	28.33 (4.74)	28.05 (4.05)	24.38 (3.20)	27.34 (5.42)	28.69 (4.52)	25.66 (4.11)	26.19 (5.72)	29.42 (5.56)	26.28 (4.95)					
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
$RMSEA_B$	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)					
CFI_B	0.34 (0.45)	0.27 (0.43)	0.06 (0.22)	0.34 (0.46)	0.50 (0.48)	0.21 (0.40)	0.46 (0.48)	0.60 (0.45)	0.39 (0.47)					
TLI_B	0.32 (0.45)	0.26 (0.43)	0.05 (0.22)	0.33 (0.46)	0.48 (0.48)	0.20 (0.40)	0.45 (0.49)	0.58 (0.46)	0.38 (0.47)					
$SRMR_B$	0.14 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.05 (0.01)					
<i>W-I-s</i>														
χ^2_W (<i>df</i> = 34)	32.36 (6.17)	31.31 (5.90)	24.47 (3.28)	33.34 (6.82)	32.61 (7.06)	26.60 (4.40)	32.78 (6.77)	33.79 (6.74)	29.24 (6.20)					
Type I error rate	0.04	0.00	0.00	0.03	0.03	0.00	0.02	0.02	0.00					
$RMSEA_W$	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)					
CFI_W	0.93 (0.16)	0.97 (0.06)	1.00 (0.00)	0.96 (0.07)	0.98 (0.04)	1.00 (0.00)	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)					
TLI_W	0.92 (0.18)	0.96 (0.08)	1.00 (0.00)	0.94 (0.09)	0.97 (0.05)	1.00 (0.00)	0.97 (0.05)	0.98 (0.02)	1.00 (0.01)					
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.00)	0.07 (0.01)	0.05 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)					
M_B														
<i>B-I-s</i>														
χ^2_B (<i>df</i> = 35)	30.15 (4.75)	29.67 (3.98)	25.71 (3.11)	29.44 (5.25)	30.48 (4.76)	27.00 (4.21)	28.41 (6.10)	31.69 (6.15)	27.95 (5.31)					
Power	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
$RMSEA_B$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)					
CFI_B	0.32 (0.42)	0.27 (0.42)	0.07 (0.23)	0.34 (0.45)	0.49 (0.48)	0.21 (0.39)	0.45 (0.47)	0.57 (0.44)	0.39 (0.47)					
TLI_B	0.28 (0.43)	0.26 (0.41)	0.06 (0.22)	0.32 (0.45)	0.47 (0.48)	0.20 (0.39)	0.43 (0.48)	0.53 (0.45)	0.38 (0.47)					
$SRMR_B$	0.15 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)					
M_W														
<i>W-I-s</i>														
χ^2_W (<i>df</i> = 35)	37.93 (6.77)	42.21 (8.46)	40.53 (8.48)	45.43 (10.12)	57.77 (13.43)	71.37 (13.80)	59.65 (12.58)	90.49 (17.58)	146.12 (23.61)					
Power	0.06	0.15	0.13	0.28	0.65	0.99	0.77	0.99	1.00					
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.00)	0.02 (0.00)					
CFI_W	0.84 (0.19)	0.82 (0.15)	0.89 (0.12)	0.80 (0.15)	0.77 (0.11)	0.78 (0.07)	0.76 (0.10)	0.74 (0.07)	0.75 (0.05)					
TLI_W	0.80 (0.22)	0.77 (0.20)	0.85 (0.15)	0.74 (0.19)	0.70 (0.15)	0.72 (0.09)	0.70 (0.13)	0.67 (0.09)	0.68 (0.06)					
$SRMR_W$	0.10 (0.01)	0.08 (0.01)	0.06 (0.01)	0.08 (0.01)	0.06 (0.01)	0.05 (0.01)	0.06 (0.01)	0.05 (0.01)	0.05 (0.00)					

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 8 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC1 (.091) for THR = 80:20 and FL = .55

	100						200		
	50			100			200		
	10	20	50	10	50	100	10	20	50
M_C	<i>B-I-s</i>								
χ^2_B (<i>df</i> = 34)	28.78 (5.04)	28.60 (4.44)	26.34 (3.76)	28.90 (5.38)	29.35 (5.44)	28.04 (4.69)	28.56 (5.75)	30.88 (6.14)	29.37 (5.97)
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
$RMSEA_B$	0.01 (0.02)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)
CFI_B	0.29 (0.43)	0.36 (0.46)	0.21 (0.40)	0.52 (0.48)	0.61 (0.46)	0.50 (0.48)	0.56 (0.47)	0.76 (0.37)	0.70 (0.43)
TLI_B	0.27 (0.43)	0.34 (0.47)	0.20 (0.39)	0.51 (0.48)	0.59 (0.46)	0.49 (0.48)	0.55 (0.47)	0.74 (0.39)	0.68 (0.43)
$SRMR_B$	0.15 (0.03)	0.12 (0.02)	0.11 (0.02)	0.10 (0.02)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)
M_B	<i>W-I-s</i>								
χ^2_W (<i>df</i> = 34)	34.08 (6.06)	33.06 (6.44)	27.19 (4.39)	34.06 (6.88)	32.97 (6.51)	29.54 (5.88)	34.05 (7.34)	33.31 (7.61)	31.69 (7.53)
Type I error rate	0.03	0.01	0.00	0.01	0.02	0.01	0.03	0.03	0.01
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
CFI_W	0.96 (0.07)	0.99 (0.03)	1.00 (0.00)	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)
TLI_W	0.95 (0.09)	0.98 (0.03)	1.00 (0.00)	0.98 (0.04)	0.99 (0.02)	1.00 (0.01)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.00)	0.06 (0.01)	0.04 (0.00)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)
M_W	<i>B-I-s</i>								
χ^2_B (<i>df</i> = 35)	30.96 (5.57)	31.44 (5.35)	28.71 (5.11)	31.69 (6.38)	32.42 (6.28)	30.45 (5.45)	31.89 (7.71)	33.74 (7.28)	32.11 (6.94)
Power	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
CFI_B	0.28 (0.39)	0.35 (0.44)	0.22 (0.39)	0.46 (0.44)	0.58 (0.44)	0.47 (0.47)	0.53 (0.45)	0.70 (0.38)	0.67 (0.42)
TLI_B	0.24 (0.39)	0.32 (0.44)	0.20 (0.39)	0.42 (0.45)	0.55 (0.46)	0.45 (0.47)	0.50 (0.46)	0.67 (0.40)	0.64 (0.43)
$SRMR_B$	0.16 (0.02)	0.13 (0.02)	0.12 (0.02)	0.11 (0.02)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)
M_W	<i>W-I-s</i>								
χ^2_W (<i>df</i> = 35)	50.92 (10.50)	71.71 (17.72)	96.36 (25.88)	77.62 (20.27)	116.33 (23.16)	211.12 (33.27)	124.85 (30.15)	225.16 (35.78)	475.38 (57.33)
Power	0.66	0.97	1.00	0.98	1.00	1.00	1.00	1.00	1.00
$RMSEA_W$	0.03 (0.01)	0.03 (0.01)	0.03 (0.00)	0.03 (0.01)	0.03 (0.00)	0.03 (0.00)	0.04 (0.01)	0.04 (0.00)	0.04 (0.00)
CFI_W	0.77 (0.13)	0.75 (0.11)	0.76 (0.07)	0.74 (0.10)	0.75 (0.07)	0.74 (0.04)	0.73 (0.06)	0.73 (0.04)	0.73 (0.03)
TLI_W	0.70 (0.17)	0.68 (0.14)	0.70 (0.08)	0.66 (0.13)	0.67 (0.09)	0.66 (0.06)	0.65 (0.08)	0.65 (0.06)	0.66 (0.04)
$SRMR_W$	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.10 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.00)

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 9 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC1 (.091) for THR = 80:20 and FL = .70

	100						200								
	50			20			50			20			50		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 34)	28.87 (4.77)	30.08 (4.55)	29.09 (4.41)	29.63 (5.40)	30.70 (5.45)	30.88 (5.45)	30.45 (6.25)	32.06 (6.87)	32.77 (6.80)						
Type I error rate	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.03	0.02						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)						
CFI_B	0.43 (0.47)	0.49 (0.47)	0.57 (0.48)	0.60 (0.45)	0.77 (0.37)	0.81 (0.34)	0.75 (0.38)	0.88 (0.24)	0.91 (0.18)						
TLL_B	0.42 (0.47)	0.47 (0.47)	0.56 (0.48)	0.58 (0.46)	0.75 (0.39)	0.79 (0.35)	0.73 (0.40)	0.85 (0.27)	0.88 (0.20)						
$SRMR_B$	0.17 (0.02)	0.14 (0.02)	0.11 (0.01)	0.12 (0.01)	0.10 (0.01)	0.08 (0.01)	0.09 (0.01)	0.07 (0.01)	0.06 (0.01)						
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 34)	34.55 (7.18)	34.03 (6.32)	31.16 (5.87)	34.22 (6.95)	33.60 (7.34)	32.17 (6.54)	34.63 (7.51)	34.47 (8.03)	33.91 (7.48)						
Type I error rate	0.04	0.03	0.01	0.02	0.05	0.01	0.03	0.06	0.03						
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)						
TLL_W	0.98 (0.04)	0.99 (0.01)	1.00 (0.00)	0.99 (0.02)	1.00 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)						
$SRMR_W$	0.08 (0.01)	0.05 (0.01)	0.03 (0.00)	0.06 (0.01)	0.04 (0.00)	0.02 (0.00)	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)						
M_B															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 35)	31.20 (4.92)	32.34 (4.60)	31.15 (4.52)	32.34 (5.83)	33.32 (5.82)	34.29 (6.05)	33.46 (6.68)	36.21 (7.55)	38.65 (8.16)						
Power	0.00	0.01	0.00	0.01	0.01	0.00	0.02	0.06	0.09						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.02)	0.02 (0.02)						
CFI_B	0.38 (0.44)	0.48 (0.45)	0.54 (0.46)	0.55 (0.44)	0.72 (0.38)	0.73 (0.35)	0.69 (0.38)	0.80 (0.27)	0.79 (0.25)						
TLL_B	0.35 (0.44)	0.46 (0.45)	0.52 (0.46)	0.52 (0.45)	0.68 (0.40)	0.69 (0.37)	0.66 (0.40)	0.75 (0.31)	0.74 (0.27)						
$SRMR_B$	0.18 (0.03)	0.14 (0.02)	0.12 (0.01)	0.12 (0.02)	0.10 (0.01)	0.09 (0.01)	0.09 (0.01)	0.07 (0.01)	0.06 (0.01)						
M_W															
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 35)	78.47 (17.14)	130.55 (30.95)	240.31 (39.21)	135.08 (29.83)	243.67 (42.03)	546.51 (64.05)	251.43 (39.14)	493.04 (59.17)	1189.24 (101.69)						
Power	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.05 (0.01)	0.05 (0.01)	0.05 (0.00)	0.05 (0.01)	0.05 (0.01)	0.05 (0.00)	0.06 (0.01)	0.06 (0.00)	0.06 (0.00)						
CFI_W	0.73 (0.09)	0.72 (0.08)	0.73 (0.04)	0.72 (0.07)	0.72 (0.05)	0.72 (0.03)	0.72 (0.04)	0.72 (0.03)	0.72 (0.02)						
TLL_W	0.65 (0.12)	0.64 (0.10)	0.65 (0.06)	0.64 (0.09)	0.65 (0.06)	0.64 (0.04)	0.64 (0.06)	0.64 (0.04)	0.64 (0.03)						
$SRMR_W$	0.13 (0.02)	0.12 (0.01)	0.11 (0.01)	0.12 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)						

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 10 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC2 (.231) for THR = 50:50 and FL= .40

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 34)	29.02 (4.82)	30.52 (4.78)	26.03 (3.56)	29.65 (5.40)	31.36 (5.68)	28.72 (5.01)	30.05 (6.29)	32.83 (6.97)	31.58 (6.54)						
Type I error rate	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01
$RMSEA_B$	0.00 (0.02)	0.01 (0.02)	0.00 (0.00)	0.01 (0.01)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
CFI_B	0.42 (0.48)	0.52 (0.45)	0.23 (0.41)	0.60 (0.45)	0.72 (0.38)	0.57 (0.47)	0.71 (0.40)	0.81 (0.30)	0.83 (0.31)						
TLI_B	0.40 (0.48)	0.48 (0.46)	0.22 (0.41)	0.57 (0.46)	0.69 (0.40)	0.56 (0.47)	0.69 (0.41)	0.77 (0.32)	0.80 (0.33)						
$SRMR_B$	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)	0.05 (0.01)						
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 34)	32.66 (5.58)	31.27 (5.43)	23.95 (3.46)	32.67 (6.36)	33.09 (7.27)	26.06 (4.83)	32.99 (7.53)	33.36 (7.32)	29.38 (6.51)						
Type I error rate	0.00	0.00	0.00	0.02	0.03	0.00	0.03	0.03	0.00						
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.96 (0.07)	0.98 (0.04)	1.00 (0.00)	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)	0.99 (0.03)	0.99 (0.01)	1.00 (0.00)						
TLI_W	0.95 (0.09)	0.98 (0.05)	1.00 (0.00)	0.97 (0.05)	0.98 (0.03)	1.00 (0.00)	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)						
$SRMR_W$	0.07 (0.01)	0.05 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)						
M_B															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 35)	30.99 (4.85)	32.43 (5.02)	27.73 (3.73)	31.74 (5.47)	33.69 (6.01)	30.97 (5.26)	32.62 (6.38)	35.96 (7.64)	34.88 (7.04)						
Power	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.04	0.02						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)						
CFI_B	0.41 (0.45)	0.49 (0.44)	0.22 (0.40)	0.58 (0.44)	0.68 (0.39)	0.55 (0.45)	0.67 (0.40)	0.75 (0.30)	0.76 (0.32)						
TLI_B	0.38 (0.45)	0.46 (0.44)	0.21 (0.40)	0.55 (0.45)	0.64 (0.41)	0.53 (0.46)	0.64 (0.41)	0.70 (0.33)	0.72 (0.34)						
$SRMR_B$	0.13 (0.01)	0.12 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)						
M_W															
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 35)	42.55 (7.76)	49.50 (10.18)	47.54 (9.04)	56.16 (12.10)	77.13 (15.48)	97.73 (17.58)	82.28 (17.40)	132.39 (22.90)	228.01 (28.25)						
Power	0.18	0.43	0.41	0.66	0.99	1.00	0.99	1.00	1.00						
$RMSEA_W$	0.02 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.03 (0.00)	0.02 (0.00)						
CFI_W	0.80 (0.16)	0.79 (0.12)	0.85 (0.09)	0.76 (0.12)	0.75 (0.08)	0.77 (0.05)	0.73 (0.08)	0.74 (0.05)	0.74 (0.03)						
TLI_W	0.74 (0.20)	0.73 (0.16)	0.81 (0.11)	0.69 (0.15)	0.67 (0.10)	0.70 (0.07)	0.66 (0.11)	0.66 (0.07)	0.67 (0.04)						
$SRMR_W$	0.08 (0.01)	0.06 (0.01)	0.05 (0.01)	0.06 (0.01)	0.05 (0.01)	0.05 (0.00)	0.06 (0.01)	0.05 (0.00)	0.05 (0.00)						

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. *B-I-s* = Between-level specific. *W-I-s* = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 11 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC2 (.231) for THR = 50:50 and FL=.55

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 34)	30.59 (4.66)	31.72 (4.97)	28.64 (4.23)	31.15 (5.83)	33.29 (6.54)	31.27 (5.33)	32.27 (7.03)	34.27 (7.20)	34.23 (7.00)						
Type I error rate	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.04	0.03						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)	0.02 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)						
CFI_B	0.60 (0.45)	0.78 (0.35)	0.71 (0.44)	0.82 (0.33)	0.89 (0.20)	0.93 (0.17)	0.92 (0.16)	0.93 (0.11)	0.94 (0.09)						
TLL_B	0.57 (0.46)	0.76 (0.37)	0.70 (0.44)	0.79 (0.34)	0.86 (0.23)	0.92 (0.19)	0.90 (0.19)	0.91 (0.14)	0.93 (0.12)						
$SRMR_B$	0.13 (0.02)	0.11 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.05 (0.01)						
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 34)	32.95 (6.05)	32.81 (6.19)	25.63 (4.00)	33.58 (6.83)	33.79 (6.98)	27.82 (5.36)	33.53 (7.67)	33.83 (7.58)	30.77 (6.99)						
Type I error rate	0.01	0.01	0.00	0.03	0.03	0.00	0.04	0.03	0.01						
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.98 (0.03)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)						
TLL_W	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)						
$SRMR_W$	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)						
M_B															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 35)	32.91 (5.04)	34.72 (5.44)	31.13 (4.36)	35.02 (6.58)	38.33 (7.49)	36.45 (6.17)	39.55 (8.28)	44.27 (9.57)	46.21 (9.77)						
Power	0.00	0.01	0.00	0.01	0.08	0.02	0.11	0.24	0.33						
$RMSEA_B$	0.01 (0.02)	0.02 (0.03)	0.01 (0.02)	0.02 (0.02)	0.03 (0.03)	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.04 (0.02)						
CFI_B	0.53 (0.44)	0.70 (0.36)	0.68 (0.43)	0.73 (0.34)	0.77 (0.25)	0.83 (0.22)	0.78 (0.22)	0.75 (0.19)	0.76 (0.17)						
TLL_B	0.49 (0.45)	0.66 (0.39)	0.66 (0.44)	0.69 (0.37)	0.71 (0.30)	0.78 (0.26)	0.72 (0.27)	0.68 (0.24)	0.69 (0.21)						
$SRMR_B$	0.13 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)						
M_W															
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 35)	66.66 (15.21)	97.73 (20.90)	125.21 (24.27)	108.28 (24.75)	188.35 (32.10)	308.93 (42.32)	194.41 (34.42)	369.19 (48.79)	759.49 (67.98)						
Power	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.04 (0.01)	0.04 (0.01)	0.03 (0.00)	0.05 (0.01)	0.05 (0.00)	0.04 (0.00)	0.05 (0.01)	0.05 (0.00)	0.05 (0.00)						
CFI_W	0.73 (0.12)	0.73 (0.07)	0.76 (0.05)	0.72 (0.08)	0.71 (0.06)	0.73 (0.03)	0.71 (0.06)	0.72 (0.04)	0.72 (0.02)						
TLL_W	0.66 (0.15)	0.66 (0.09)	0.69 (0.06)	0.64 (0.11)	0.63 (0.07)	0.66 (0.04)	0.63 (0.07)	0.64 (0.05)	0.65 (0.03)						
$SRMR_W$	0.10 (0.01)	0.08 (0.01)	0.08 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.00)	0.08 (0.01)	0.08 (0.00)	0.07 (0.00)						

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-I-s = Between-level specific. W-I-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 12 Means of level-specific χ^2 test statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC2 (.231) for THR = 50:50 and FL=.70

	100						200		
	50			100			200		
	10	20	50	10	20	50	10	20	50
M_C									
<i>B-l-s</i>									
χ^2_B (<i>df</i> = 34)	32.32 (4.89)	32.80 (5.52)	31.17 (5.25)	32.87 (5.94)	34.14 (6.85)	33.71 (6.59)	33.97 (7.66)	34.74 (7.48)	35.39 (7.70)
Type I error rate	0.01	0.01	0.00	0.01	0.03	0.03	0.04	0.04	0.06
$RMSEA_B$	0.01 (0.02)	0.02 (0.03)	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)
CFI_B	0.86 (0.27)	0.90 (0.20)	0.95 (0.14)	0.93 (0.13)	0.94 (0.10)	0.96 (0.07)	0.96 (0.07)	0.97 (0.05)	0.97 (0.04)
TLL_B	0.83 (0.30)	0.87 (0.24)	0.93 (0.16)	0.92 (0.16)	0.92 (0.13)	0.95 (0.10)	0.95 (0.09)	0.96 (0.06)	0.96 (0.05)
$SRMR_B$	0.14 (0.02)	0.12 (0.02)	0.11 (0.01)	0.10 (0.01)	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)	0.05 (0.01)
<i>W-l-s</i>									
χ^2_W (<i>df</i> = 34)	33.95 (6.68)	34.33 (6.21)	27.93 (4.78)	33.81 (6.79)	34.78 (7.50)	29.85 (6.13)	33.73 (7.85)	34.36 (7.97)	32.17 (7.41)
Type I error rate	0.03	0.02	0.00	0.03	0.05	0.01	0.04	0.05	0.03
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
CFI_W	0.99 (0.02)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
TLL_W	1.00 (0.02)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)
$SRMR_W$	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)
M_B									
<i>B-l-s</i>									
χ^2_B (<i>df</i> = 35)	36.14 (5.48)	39.19 (7.01)	38.11 (6.80)	42.87 (8.63)	49.80 (10.56)	52.81 (11.40)	57.63 (13.31)	71.52 (16.58)	84.13 (18.62)
Power	0.02	0.07	0.06	0.19	0.45	0.56	0.70	0.92	0.98
$RMSEA_B$	0.03 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.06 (0.03)	0.07 (0.03)	0.05 (0.02)	0.07 (0.02)	0.08 (0.02)
CFI_B	0.75 (0.30)	0.72 (0.27)	0.80 (0.23)	0.72 (0.23)	0.68 (0.19)	0.70 (0.17)	0.67 (0.17)	0.65 (0.13)	0.64 (0.12)
TLL_B	0.69 (0.35)	0.65 (0.33)	0.74 (0.28)	0.64 (0.29)	0.59 (0.24)	0.61 (0.22)	0.58 (0.22)	0.55 (0.17)	0.54 (0.16)
$SRMR_B$	0.15 (0.02)	0.13 (0.02)	0.12 (0.02)	0.12 (0.02)	0.11 (0.01)	0.10 (0.01)	0.10 (0.01)	0.10 (0.01)	0.09 (0.01)
M_W									
<i>W-l-s</i>									
χ^2_W (<i>df</i> = 35)	114.23 (24.71)	199.21 (37.19)	333.31 (50.75)	212.85 (40.29)	409.32 (54.57)	838.05 (85.36)	414.50 (59.55)	838.08 (81.56)	1946.94 (134.47)
Power	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$RMSEA_W$	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.00)	0.07 (0.01)	0.08 (0.00)	0.07 (0.00)
CFI_W	0.71 (0.09)	0.71 (0.06)	0.73 (0.04)	0.70 (0.06)	0.70 (0.04)	0.71 (0.03)	0.70 (0.05)	0.70 (0.03)	0.71 (0.02)
TLL_W	0.62 (0.11)	0.62 (0.08)	0.65 (0.05)	0.62 (0.08)	0.61 (0.05)	0.63 (0.04)	0.61 (0.06)	0.62 (0.04)	0.62 (0.02)
$SRMR_W$	0.12 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-l-s = Between-level specific. W-l-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 test statistics are reported to two decimal places for the sake of space

Table 13 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC2 (.231) for THR = 80:20 and FL=.40

	100						200					
	50		20		50		10		20		50	
	10	50	10	50	10	50	10	50	10	50	10	50
M_C												
<i>B-l-s</i>												
χ^2_B (<i>df</i> = 34)	28.23 (4.64)	28.33 (4.10)	24.40 (2.99)	29.16 (5.38)	29.88 (5.09)	26.39 (4.02)	29.44 (6.66)	31.84 (6.15)	28.81 (5.80)			
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00			
$RMSEA_B$	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)			
CFI_B	0.39 (0.46)	0.38 (0.47)	0.09 (0.27)	0.54 (0.47)	0.66 (0.44)	0.37 (0.48)	0.65 (0.44)	0.82 (0.31)	0.72 (0.42)			
TLL_B	0.37 (0.46)	0.36 (0.47)	0.08 (0.28)	0.53 (0.47)	0.64 (0.44)	0.37 (0.48)	0.63 (0.44)	0.79 (0.33)	0.71 (0.42)			
$SRMR_B$	0.14 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)			
<i>W-l-s</i>												
χ^2_W (<i>df</i> = 34)	32.76 (5.82)	31.10 (5.46)	24.39 (3.35)	32.76 (6.72)	32.41 (6.47)	26.36 (4.86)	32.74 (7.27)	33.88 (7.38)	29.25 (5.84)			
Type I error rate	0.02	0.00	0.00	0.02	0.02	0.00	0.02	0.03	0.00			
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)			
CFI_W	0.92 (0.17)	0.97 (0.10)	1.00 (0.00)	0.96 (0.07)	0.98 (0.04)	1.00 (0.01)	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)			
TLL_W	0.90 (0.19)	0.96 (0.11)	1.00 (0.00)	0.95 (0.09)	0.98 (0.05)	1.00 (0.01)	0.97 (0.05)	0.98 (0.03)	1.00 (0.01)			
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.01)	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)			
M_B												
<i>B-l-s</i>												
χ^2_B (<i>df</i> = 35)	30.18 (4.77)	30.37 (4.14)	25.91 (3.04)	31.57 (5.81)	32.07 (5.37)	28.31 (4.19)	32.12 (7.16)	34.78 (6.72)	31.79 (6.61)			
Power	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.02			
$RMSEA_B$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.02)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)			
CFI_B	0.38 (0.45)	0.37 (0.45)	0.09 (0.27)	0.49 (0.45)	0.64 (0.43)	0.37 (0.47)	0.59 (0.43)	0.76 (0.32)	0.68 (0.41)			
TLL_B	0.36 (0.45)	0.35 (0.45)	0.08 (0.28)	0.46 (0.46)	0.61 (0.43)	0.36 (0.47)	0.56 (0.44)	0.72 (0.35)	0.65 (0.42)			
$SRMR_B$	0.15 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)			
M_W												
<i>W-l-s</i>												
χ^2_W (<i>df</i> = 35)	37.42 (6.23)	41.88 (9.05)	39.55 (7.21)	45.10 (9.71)	56.53 (12.91)	70.90 (14.06)	58.29 (12.96)	90.99 (18.79)	144.50 (23.95)			
Power	0.02	0.18	0.09	0.30	0.65	0.95	0.71	0.99	1.00			
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.00)	0.02 (0.00)			
CFI_W	0.84 (0.19)	0.82 (0.16)	0.90 (0.10)	0.80 (0.16)	0.78 (0.11)	0.78 (0.07)	0.78 (0.11)	0.75 (0.07)	0.75 (0.05)			
TLL_W	0.80 (0.23)	0.77 (0.20)	0.87 (0.13)	0.74 (0.20)	0.72 (0.14)	0.72 (0.09)	0.71 (0.14)	0.67 (0.09)	0.68 (0.06)			
$SRMR_W$	0.10 (0.01)	0.08 (0.01)	0.06 (0.01)	0.08 (0.01)	0.06 (0.01)	0.05 (0.01)	0.06 (0.01)	0.05 (0.01)	0.05 (0.00)			

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-l-s = Between-level specific, W-l-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 14 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC2 (.231) for THR = 80:20 and FL=.55

	100						200		
	50			100			200		
	10	20	50	10	50	100	10	20	50
M_C	<i>B-I-s</i>								
χ^2_B (<i>df</i> = 34)	29.65 (4.76)	30.35 (4.89)	26.87 (3.94)	30.64 (5.72)	31.70 (5.50)	29.38 (5.11)	31.75 (6.77)	33.07 (6.66)	32.16 (6.52)
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.02
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
CFI_B	0.55 (0.46)	0.61 (0.45)	0.50 (0.49)	0.74 (0.39)	0.86 (0.28)	0.91 (0.25)	0.89 (0.22)	0.93 (0.14)	0.95 (0.10)
TLI_B	0.53 (0.47)	0.59 (0.46)	0.49 (0.49)	0.72 (0.40)	0.84 (0.29)	0.90 (0.26)	0.87 (0.24)	0.91 (0.16)	0.94 (0.12)
$SRMR_B$	0.15 (0.03)	0.12 (0.02)	0.11 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.06 (0.01)	0.06 (0.01)
M_C	<i>W-I-s</i>								
χ^2_W (<i>df</i> = 34)	33.73 (6.13)	32.36 (5.79)	26.50 (4.05)	33.86 (7.11)	33.24 (6.55)	28.90 (5.63)	33.84 (7.55)	33.92 (7.72)	31.41 (6.97)
Type I error rate	0.03	0.01	0.00	0.03	0.02	0.01	0.05	0.05	0.02
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
CFI_W	0.97 (0.06)	0.99 (0.02)	1.00 (0.00)	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)
TLI_W	0.96 (0.08)	0.99 (0.03)	1.00 (0.00)	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.00)	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)
M_B	<i>B-I-s</i>								
χ^2_B (<i>df</i> = 35)	32.22 (5.18)	33.28 (5.58)	29.29 (4.36)	33.80 (6.21)	36.13 (6.51)	33.79 (6.16)	37.35 (7.87)	41.48 (8.66)	42.28 (8.82)
Power	0.00	0.00	0.00	0.02	0.04	0.03	0.10	0.26	0.37
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)	0.02 (0.02)	0.01 (0.02)	0.02 (0.02)	0.03 (0.02)	0.03 (0.02)
CFI_B	0.51 (0.45)	0.55 (0.44)	0.49 (0.49)	0.66 (0.39)	0.75 (0.30)	0.83 (0.27)	0.79 (0.26)	0.78 (0.20)	0.79 (0.18)
TLI_B	0.48 (0.46)	0.52 (0.45)	0.49 (0.49)	0.62 (0.41)	0.70 (0.33)	0.80 (0.30)	0.74 (0.30)	0.72 (0.25)	0.73 (0.23)
$SRMR_B$	0.16 (0.02)	0.13 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)
M_W	<i>W-I-s</i>								
χ^2_W (<i>df</i> = 35)	52.38 (13.66)	69.88 (17.82)	87.57 (17.16)	75.23 (16.85)	117.25 (24.56)	200.70 (33.39)	123.56 (24.19)	222.66 (34.94)	459.52 (54.41)
Power	0.67	0.95	1.00	0.99	1.00	1.00	1.00	1.00	1.00
$RMSEA_W$	0.03 (0.01)	0.03 (0.01)	0.02 (0.00)	0.03 (0.01)	0.03 (0.01)	0.03 (0.00)	0.04 (0.00)	0.04 (0.00)	0.03 (0.00)
CFI_W	0.76 (0.14)	0.75 (0.11)	0.77 (0.06)	0.74 (0.09)	0.74 (0.07)	0.74 (0.04)	0.73 (0.06)	0.73 (0.04)	0.73 (0.03)
TLI_W	0.69 (0.18)	0.68 (0.14)	0.70 (0.08)	0.66 (0.12)	0.67 (0.08)	0.67 (0.06)	0.65 (0.08)	0.65 (0.06)	0.66 (0.04)
$SRMR_W$	0.11 (0.02)	0.09 (0.01)	0.08 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.00)

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 15 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC2 (.231) for THR = 80:20 and FL=.70

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 34)	29.85 (4.66)	30.46 (4.79)	27.74 (4.10)	31.24 (5.63)	31.83 (5.56)	30.16 (5.23)	32.30 (6.76)	33.45 (6.73)	32.51 (6.78)						
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.02						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)						
CFI_B	0.57 (0.46)	0.68 (0.43)	0.64 (0.47)	0.83 (0.30)	0.90 (0.22)	0.93 (0.20)	0.92 (0.16)	0.94 (0.10)	0.96 (0.07)						
TLI_B	0.55 (0.47)	0.66 (0.43)	0.63 (0.47)	0.80 (0.32)	0.88 (0.24)	0.92 (0.21)	0.91 (0.18)	0.93 (0.12)	0.95 (0.09)						
$SRMR_B$	0.15 (0.02)	0.13 (0.02)	0.11 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.06 (0.01)	0.06 (0.01)						
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 34)	34.43 (6.79)	32.04 (5.72)	27.08 (4.48)	33.82 (7.05)	33.05 (6.56)	29.28 (6.03)	34.09 (7.64)	33.89 (7.74)	31.53 (6.78)						
Type I error rate	0.03	0.01	0.00	0.03	0.02	0.01	0.05	0.05	0.02						
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.97 (0.05)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)						
TLI_W	0.96 (0.07)	0.99 (0.02)	1.00 (0.00)	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.02)	1.00 (0.01)	1.00 (0.00)						
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.00)	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)						
M_B															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 35)	32.30 (4.71)	33.47 (4.85)	30.46 (4.36)	35.04 (6.09)	37.15 (6.42)	36.02 (6.64)	39.33 (8.09)	44.56 (9.03)	47.25 (10.02)						
Power	0.00	0.00	0.00	0.02	0.04	0.03	0.10	0.26	0.37						
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.04 (0.02)						
CFI_B	0.54 (0.45)	0.63 (0.42)	0.61 (0.46)	0.73 (0.33)	0.78 (0.26)	0.83 (0.24)	0.79 (0.22)	0.75 (0.18)	0.75 (0.16)						
TLI_B	0.51 (0.46)	0.60 (0.43)	0.60 (0.47)	0.68 (0.36)	0.72 (0.30)	0.79 (0.28)	0.74 (0.26)	0.68 (0.23)	0.68 (0.20)						
$SRMR_B$	0.16 (0.02)	0.13 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)						
M_W															
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 35)	56.10 (12.99)	78.50 (19.17)	107.38 (21.30)	84.71 (19.43)	137.36 (28.03)	248.06 (37.75)	143.65 (28.69)	265.37 (38.75)	565.15 (61.78)						
Power	0.67	0.95	1.00	0.99	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.03 (0.01)	0.03 (0.01)	0.03 (0.00)	0.04 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.04 (0.00)	0.04 (0.00)						
CFI_W	0.76 (0.13)	0.75 (0.10)	0.76 (0.06)	0.73 (0.09)	0.74 (0.06)	0.74 (0.04)	0.73 (0.06)	0.73 (0.04)	0.73 (0.03)						
TLI_W	0.69 (0.16)	0.67 (0.12)	0.69 (0.07)	0.66 (0.12)	0.66 (0.08)	0.66 (0.05)	0.65 (0.08)	0.65 (0.05)	0.65 (0.03)						
$SRMR_W$	0.12 (0.01)	0.10 (0.01)	0.09 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.00)						

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-I-s = Between-level specific. W-I-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 16 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC3 (.333) for THR = 50:50 and FL= .40

	50			100			200		
	NG =			GS =			FL =		
	10	20	50	10	20	50	10	20	50
M_C									
<i>B-I-s</i>									
χ^2_B (<i>df</i> = 34)	29.82 (4.41)	31.19 (4.65)	26.51 (3.54)	30.64 (5.83)	32.72 (6.18)	29.79 (5.35)	31.60 (6.45)	33.94 (6.96)	32.78 (6.78)
Type I error rate	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.02
$RMSEA_B$	0.01 (0.01)	0.01 (0.02)	0.00 (0.00)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)
CFI_B	0.53 (0.47)	0.65 (0.43)	0.41 (0.49)	0.73 (0.39)	0.84 (0.28)	0.81 (0.36)	0.89 (0.22)	0.90 (0.17)	0.91 (0.18)
TLI_B	0.51 (0.48)	0.63 (0.43)	0.40 (0.49)	0.71 (0.41)	0.81 (0.30)	0.79 (0.36)	0.87 (0.24)	0.88 (0.20)	0.89 (0.20)
$SRMR_B$	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.06 (0.01)	0.06 (0.01)	0.05 (0.01)
<i>W-I-s</i>									
χ^2_W (<i>df</i> = 34)	32.51 (5.63)	30.86 (5.67)	23.74 (3.29)	33.05 (6.46)	32.83 (7.33)	25.82 (4.55)	33.47 (7.38)	33.74 (7.63)	29.22 (6.32)
Type I error rate	0.00	0.01	0.00	0.02	0.02	0.00	0.04	0.03	0.00
$RMSEA_W$	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
CFI_W	0.96 (0.07)	0.99 (0.03)	1.00 (0.00)	0.98 (0.04)	0.99 (0.03)	1.00 (0.00)	0.99 (0.03)	0.99 (0.01)	1.00 (0.00)
TLI_W	0.95 (0.09)	0.98 (0.04)	1.00 (0.00)	0.97 (0.06)	0.98 (0.03)	1.00 (0.00)	0.98 (0.03)	0.99 (0.02)	1.00 (0.00)
$SRMR_W$	0.07 (0.01)	0.05 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)
M_B									
<i>B-I-s</i>									
χ^2_B (<i>df</i> = 35)	32.00 (4.61)	33.81 (5.27)	28.51 (3.72)	33.75 (6.17)	36.10 (6.70)	33.07 (5.94)	36.68 (7.02)	40.15 (8.52)	39.64 (8.09)
Power	0.00	0.01	0.00	0.01	0.04	0.01	0.04	0.13	0.11
$RMSEA_B$	0.01 (0.02)	0.02 (0.03)	0.00 (0.01)	0.01 (0.02)	0.02 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
CFI_B	0.48 (0.45)	0.57 (0.43)	0.39 (0.48)	0.68 (0.39)	0.77 (0.30)	0.75 (0.35)	0.78 (0.26)	0.77 (0.22)	0.79 (0.22)
TLI_B	0.45 (0.45)	0.53 (0.44)	0.39 (0.48)	0.64 (0.41)	0.72 (0.33)	0.72 (0.37)	0.73 (0.31)	0.71 (0.27)	0.73 (0.26)
$SRMR_B$	0.13 (0.01)	0.12 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)
M_W									
<i>W-I-s</i>									
χ^2_W (<i>df</i> = 35)	42.69 (7.97)	48.59 (10.22)	46.20 (8.64)	56.39 (12.47)	75.48 (15.96)	94.39 (16.45)	81.56 (17.33)	131.15 (23.03)	219.64 (29.05)
Power	0.17	0.41	0.34	0.68	0.97	1.00	0.98	1.00	1.00
$RMSEA_W$	0.02 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.03 (0.00)	0.02 (0.00)
CFI_W	0.80 (0.16)	0.80 (0.12)	0.86 (0.08)	0.75 (0.13)	0.75 (0.08)	0.77 (0.05)	0.74 (0.09)	0.74 (0.05)	0.74 (0.03)
TLI_W	0.74 (0.20)	0.75 (0.16)	0.82 (0.11)	0.68 (0.16)	0.68 (0.11)	0.71 (0.07)	0.66 (0.11)	0.66 (0.07)	0.67 (0.04)
$SRMR_W$	0.08 (0.01)	0.06 (0.01)	0.05 (0.01)	0.07 (0.01)	0.05 (0.01)	0.05 (0.00)	0.06 (0.01)	0.05 (0.00)	0.05 (0.00)

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 17 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC3 (.333) for THR = 50:50 and FL= .55

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 34)	31.85 (4.84)	32.58 (5.22)	28.84 (4.63)	32.44 (5.97)	34.23 (6.72)	31.88 (5.90)	33.41 (7.47)	35.01 (7.35)	34.49 (7.29)						
Type I error rate	0.00	0.01	0.00	0.01	0.03	0.01	0.04	0.06	0.04						
$RMSEA_B$	0.01 (0.02)	0.02 (0.02)	0.01 (0.01)	0.01 (0.02)	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)						
CFI_B	0.79 (0.35)	0.89 (0.23)	0.86 (0.32)	0.93 (0.12)	0.93 (0.11)	0.96 (0.09)	0.96 (0.08)	0.96 (0.06)	0.97 (0.05)						
TLL_B	0.77 (0.36)	0.86 (0.25)	0.85 (0.33)	0.91 (0.15)	0.91 (0.15)	0.95 (0.11)	0.94 (0.10)	0.95 (0.08)	0.96 (0.06)						
$SRMR_B$	0.13 (0.02)	0.11 (0.01)	0.11 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)	0.05 (0.01)						
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 34)	33.04 (6.00)	32.41 (6.06)	25.00 (3.79)	32.99 (6.37)	33.55 (7.24)	27.40 (5.19)	33.80 (7.42)	33.82 (7.57)	30.54 (7.04)						
Type I error rate	0.00	0.02	0.00	0.02	0.04	0.00	0.03	0.04	0.01						
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)						
TLL_W	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)						
$SRMR_W$	0.06 (0.01)	0.05 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)						
M_B															
<i>B-I-s</i>															
χ^2_B (<i>df</i> = 35)	35.37 (5.37)	37.68 (6.21)	32.94 (5.28)	40.45 (8.02)	45.17 (9.12)	43.17 (8.55)	51.36 (11.33)	60.31 (13.82)	64.99 (14.75)						
Power	0.01	0.05	0.00	0.13	0.28	0.22	0.52	0.77	0.86						
$RMSEA_B$	0.02 (0.03)	0.04 (0.03)	0.01 (0.02)	0.03 (0.03)	0.05 (0.03)	0.04 (0.03)	0.04 (0.02)	0.06 (0.02)	0.06 (0.02)						
CFI_B	0.69 (0.35)	0.72 (0.29)	0.80 (0.33)	0.76 (0.22)	0.72 (0.20)	0.77 (0.18)	0.71 (0.17)	0.69 (0.14)	0.69 (0.13)						
TLL_B	0.64 (0.38)	0.65 (0.33)	0.78 (0.35)	0.69 (0.28)	0.64 (0.25)	0.71 (0.23)	0.63 (0.22)	0.60 (0.18)	0.60 (0.17)						
$SRMR_B$	0.14 (0.02)	0.12 (0.02)	0.12 (0.01)	0.10 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)						
M_W															
<i>W-I-s</i>															
χ^2_W (<i>df</i> = 35)	65.32 (14.00)	95.57 (19.49)	113.50 (21.17)	106.30 (23.56)	184.19 (31.24)	290.04 (41.70)	191.20 (33.48)	360.79 (47.49)	724.92 (70.08)						
Power	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.04 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.05 (0.00)	0.04 (0.00)	0.05 (0.01)	0.05 (0.00)	0.04 (0.00)						
CFI_W	0.74 (0.11)	0.73 (0.08)	0.77 (0.05)	0.72 (0.08)	0.72 (0.05)	0.74 (0.03)	0.71 (0.06)	0.72 (0.04)	0.73 (0.02)						
TLL_W	0.66 (0.14)	0.65 (0.10)	0.70 (0.06)	0.64 (0.10)	0.63 (0.07)	0.66 (0.04)	0.63 (0.07)	0.64 (0.05)	0.65 (0.03)						
$SRMR_W$	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.00)	0.08 (0.01)	0.08 (0.00)	0.07 (0.00)						

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-I-s = Between-level specific. W-I-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 18 Means of level-specific χ^2 test statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC3 (.333) for THR = 50:50 and FL=.70.

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
$B \rightarrow I \rightarrow S$															
$\chi^2_B (df = 34)$	33.12 (5.26)	33.59 (5.79)	30.40 (5.37)	33.81 (6.57)	34.49 (7.16)	33.15 (6.72)	34.02 (7.59)	35.08 (7.74)	34.90 (7.74)						
Type I error rate	0.00	0.02	0.00	0.02	0.04	0.03	0.04	0.05	0.05						
$RMSEA_B$	0.02 (0.03)	0.02 (0.03)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)						
CFI_B	0.92 (0.15)	0.94 (0.10)	0.98 (0.06)	0.96 (0.07)	0.97 (0.05)	0.98 (0.04)	0.98 (0.04)	0.98 (0.03)	0.99 (0.02)						
TLL_B	0.90 (0.18)	0.92 (0.14)	0.97 (0.08)	0.95 (0.09)	0.95 (0.07)	0.97 (0.06)	0.97 (0.05)	0.98 (0.04)	0.98 (0.03)						
$SRMR_B$	0.13 (0.02)	0.11 (0.02)	0.10 (0.02)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)	0.05 (0.01)						
$W \rightarrow I \rightarrow S$															
$\chi^2_W (df = 34)$	33.45 (6.30)	33.48 (6.15)	26.88 (4.39)	33.68 (7.16)	34.31 (7.67)	28.91 (5.83)	33.81 (7.55)	34.66 (7.79)	31.70 (7.60)						
Type I error rate	0.01	0.02	0.00	0.03	0.05	0.00	0.03	0.05	0.02						
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)						
TLL_W	0.99 (0.02)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)						
$SRMR_W$	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)						
M_B															
$B \rightarrow I \rightarrow S$															
$\chi^2_B (df = 35)$	41.99 (7.40)	47.86 (9.33)	45.29 (8.97)	57.89 (12.50)	71.68 (15.33)	76.72 (16.22)	92.62 (20.21)	121.60 (24.40)	145.17 (26.93)						
Power	0.14	0.36	0.29	0.71	0.93	0.96	0.99	1.00	1.00						
$RMSEA_B$	0.06 (0.04)	0.08 (0.04)	0.07 (0.04)	0.08 (0.03)	0.10 (0.02)	0.11 (0.02)	0.09 (0.02)	0.11 (0.02)	0.12 (0.02)						
CFI_B	0.68 (0.25)	0.62 (0.23)	0.72 (0.20)	0.62 (0.19)	0.57 (0.16)	0.60 (0.15)	0.58 (0.14)	0.55 (0.12)	0.53 (0.11)						
TLL_B	0.60 (0.31)	0.52 (0.29)	0.64 (0.25)	0.51 (0.24)	0.44 (0.21)	0.48 (0.19)	0.46 (0.18)	0.42 (0.15)	0.40 (0.14)						
$SRMR_B$	0.16 (0.02)	0.14 (0.02)	0.13 (0.02)	0.13 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.02)	0.11 (0.02)	0.11 (0.01)						
M_W															
$W \rightarrow I \rightarrow S$															
$\chi^2_W (df = 35)$	110.56 (25.08)	190.52 (36.43)	293.23 (48.70)	207.29 (40.82)	395.82 (53.69)	760.89 (81.90)	403.04 (58.61)	809.51 (80.84)	1826.21 (133.29)						
Power	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.07 (0.01)	0.07 (0.01)	0.05 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.00)	0.07 (0.01)	0.07 (0.00)	0.07 (0.00)						
CFI_W	0.71 (0.09)	0.71 (0.06)	0.74 (0.04)	0.70 (0.07)	0.70 (0.04)	0.72 (0.03)	0.70 (0.05)	0.70 (0.03)	0.71 (0.02)						
TLL_W	0.63 (0.11)	0.63 (0.08)	0.66 (0.05)	0.62 (0.08)	0.62 (0.06)	0.64 (0.04)	0.61 (0.06)	0.62 (0.04)	0.63 (0.03)						
$SRMR_W$	0.12 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)						

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-I-s = Between-level specific. W-I-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 test statistics are reported to two decimal places for the sake of space.

Table 19 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC3 (.333) for THR = 80:20 and FL=.40

	50			100			200		
	NG =			GS =			FL =		
	10	20	50	10	20	50	10	20	50
M_C									
<i>B-l-s</i>									
χ^2_B (df = 34)	29.12 (4.65)	29.34 (4.40)	24.68 (3.20)	29.79 (5.32)	31.15 (5.33)	27.16 (4.23)	30.95 (6.60)	33.16 (6.54)	29.94 (5.98)
Type I error rate	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00
$RMSEA_B$	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
CFI_B	0.46 (0.47)	0.51 (0.47)	0.16 (0.36)	0.68 (0.43)	0.80 (0.34)	0.64 (0.47)	0.84 (0.30)	0.89 (0.21)	0.93 (0.18)
TLL_B	0.44 (0.47)	0.49 (0.48)	0.16 (0.36)	0.66 (0.44)	0.78 (0.35)	0.64 (0.47)	0.82 (0.31)	0.86 (0.24)	0.92 (0.20)
$SRMR_B$	0.14 (0.02)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)
<i>W-l-s</i>									
χ^2_W (df = 34)	32.60 (5.68)	30.79 (5.67)	23.96 (3.19)	32.93 (7.29)	32.16 (6.31)	25.84 (4.68)	32.85 (6.92)	33.34 (7.30)	28.63 (6.00)
Type I error rate	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.04	0.00
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
CFI_W	0.93 (0.15)	0.97 (0.06)	1.00 (0.00)	0.96 (0.08)	0.98 (0.04)	1.00 (0.01)	0.98 (0.04)	0.99 (0.02)	1.00 (0.01)
TLL_W	0.91 (0.18)	0.97 (0.08)	1.00 (0.00)	0.94 (0.11)	0.98 (0.05)	1.00 (0.01)	0.97 (0.05)	0.99 (0.03)	1.00 (0.01)
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.01)	0.06 (0.01)	0.05 (0.01)	0.03 (0.00)	0.05 (0.01)	0.03 (0.00)	0.02 (0.00)
M_B									
<i>B-l-s</i>									
χ^2_B (df = 35)	31.12 (4.59)	31.65 (4.66)	26.43 (3.33)	32.41 (5.67)	34.37 (5.82)	29.78 (4.74)	34.81 (7.45)	38.53 (7.90)	35.51 (7.19)
Power	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.09	0.04
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.00)	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)	0.02 (0.02)	0.01 (0.02)
CFI_B	0.45 (0.45)	0.48 (0.46)	0.17 (0.36)	0.63 (0.42)	0.72 (0.34)	0.62 (0.46)	0.77 (0.31)	0.77 (0.25)	0.85 (0.22)
TLL_B	0.42 (0.46)	0.46 (0.47)	0.16 (0.36)	0.60 (0.43)	0.68 (0.37)	0.61 (0.46)	0.73 (0.34)	0.72 (0.29)	0.82 (0.26)
$SRMR_B$	0.15 (0.02)	0.13 (0.01)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)
M_W									
<i>W-l-s</i>									
χ^2_W (df = 35)	37.35 (6.98)	40.86 (8.61)	38.14 (6.97)	45.22 (9.58)	56.25 (12.84)	68.19 (13.12)	59.73 (12.64)	88.90 (17.76)	141.92 (24.26)
Power	0.04	0.14	0.08	0.31	0.66	0.92	0.79	0.99	1.00
$RMSEA_W$	0.01 (0.01)	0.01 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)	0.02 (0.00)	0.02 (0.00)
CFI_W	0.85 (0.18)	0.84 (0.15)	0.92 (0.10)	0.80 (0.16)	0.78 (0.11)	0.79 (0.07)	0.76 (0.10)	0.75 (0.07)	0.76 (0.05)
TLL_W	0.81 (0.22)	0.80 (0.19)	0.89 (0.13)	0.74 (0.20)	0.72 (0.15)	0.73 (0.09)	0.70 (0.13)	0.68 (0.09)	0.69 (0.06)
$SRMR_W$	0.10 (0.01)	0.08 (0.01)	0.06 (0.01)	0.08 (0.01)	0.06 (0.01)	0.05 (0.01)	0.06 (0.01)	0.05 (0.01)	0.05 (0.00)

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-l-s = Between-level specific. W-l-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 20 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC3 (.333) for THR = 80:20 and FL=.55

	100						200		
	50			100			200		
	10	20	50	10	50	100	10	20	50
M_C	<i>B-I-s</i>								
χ^2_B (<i>df</i> = 34)	30.72 (4.76)	31.13 (4.89)	27.20 (3.94)	32.25 (6.09)	32.74 (5.93)	30.01 (5.36)	33.14 (6.91)	34.16 (7.00)	32.36 (6.71)
Type I error rate	0.00	0.00	0.00	0.01	0.00	0.00	0.04	0.04	0.02
$RMSEA_B$	0.01 (0.02)	0.01 (0.02)	0.00 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)
CFI_B	0.73 (0.40)	0.84 (0.31)	0.77 (0.41)	0.89 (0.21)	0.94 (0.14)	0.97 (0.09)	0.95 (0.09)	0.96 (0.07)	0.98 (0.05)
TLI_B	0.71 (0.41)	0.82 (0.33)	0.77 (0.41)	0.86 (0.24)	0.92 (0.16)	0.96 (0.10)	0.93 (0.12)	0.95 (0.09)	0.97 (0.06)
$SRMR_B$	0.15 (0.02)	0.12 (0.02)	0.11 (0.01)	0.10 (0.01)	0.09 (0.01)	0.08 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)
M_C	<i>W-I-s</i>								
χ^2_W (<i>df</i> = 34)	33.92 (6.18)	32.15 (5.65)	26.21 (4.11)	33.74 (6.89)	32.55 (6.53)	28.07 (5.46)	34.06 (7.54)	33.44 (7.64)	30.69 (6.67)
Type I error rate	0.02	0.01	0.00	0.02	0.02	0.00	0.05	0.04	0.01
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
CFI_W	0.97 (0.06)	0.99 (0.02)	1.00 (0.00)	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)
TLI_W	0.96 (0.07)	0.99 (0.03)	1.00 (0.00)	0.98 (0.04)	0.99 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.01)	1.00 (0.00)
$SRMR_W$	0.09 (0.01)	0.06 (0.01)	0.04 (0.00)	0.06 (0.01)	0.04 (0.01)	0.03 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)
M_B	<i>B-I-s</i>								
χ^2_B (<i>df</i> = 35)	34.27 (5.93)	35.58 (5.72)	30.83 (4.77)	38.29 (7.13)	41.78 (7.80)	39.69 (8.04)	46.97 (10.06)	54.94 (12.49)	57.88 (12.88)
Power	0.00	0.01	0.00	0.12	0.27	0.23	0.54	0.84	0.90
$RMSEA_B$	0.02 (0.03)	0.02 (0.03)	0.01 (0.02)	0.03 (0.02)	0.04 (0.03)	0.03 (0.03)	0.04 (0.02)	0.05 (0.02)	0.05 (0.02)
CFI_B	0.63 (0.40)	0.72 (0.34)	0.74 (0.41)	0.74 (0.27)	0.74 (0.22)	0.82 (0.19)	0.73 (0.19)	0.70 (0.15)	0.71 (0.14)
TLI_B	0.59 (0.42)	0.67 (0.37)	0.73 (0.41)	0.68 (0.32)	0.67 (0.27)	0.77 (0.24)	0.66 (0.24)	0.62 (0.20)	0.63 (0.18)
$SRMR_B$	0.16 (0.02)	0.13 (0.02)	0.12 (0.02)	0.12 (0.01)	0.10 (0.01)	0.09 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)
M_W	<i>W-I-s</i>								
χ^2_W (<i>df</i> = 35)	52.44 (13.58)	68.65 (15.93)	82.67 (17.04)	75.35 (16.44)	117.13 (25.08)	190.82 (31.84)	123.05 (23.73)	222.04 (35.39)	447.00 (52.44)
Power	0.64	0.94	1.00	0.99	1.00	1.00	1.00	1.00	1.00
$RMSEA_W$	0.03 (0.01)	0.03 (0.01)	0.02 (0.00)	0.03 (0.01)	0.03 (0.01)	0.03 (0.00)	0.04 (0.00)	0.04 (0.00)	0.03 (0.00)
CFI_W	0.76 (0.15)	0.76 (0.10)	0.78 (0.06)	0.74 (0.09)	0.74 (0.07)	0.74 (0.04)	0.73 (0.06)	0.73 (0.04)	0.74 (0.03)
TLI_W	0.69 (0.19)	0.69 (0.13)	0.71 (0.08)	0.66 (0.12)	0.67 (0.08)	0.67 (0.05)	0.66 (0.08)	0.66 (0.06)	0.66 (0.04)
$SRMR_W$	0.11 (0.02)	0.09 (0.01)	0.08 (0.01)	0.09 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.00)

Note. M_C = Correctly specified hypothesized model, M_B = Misspecification in between-level model only, M_W = Misspecification in within-level model only. B-I-s = Between-level specific, W-I-s = within-level specific, NG = Numbers of groups, GS = Group size, ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Table 21 Means of level-specific χ^2 statistics and fit indices for the M_C , M_B , and M_W within a condition of ICC3 (.333) for THR = 80:20 and FL=.70

	100						200								
	50			100			50			100			200		
	10	20	50	10	20	50	10	20	50	10	20	50	10	20	50
M_C															
<i>B-l-s</i>															
χ^2_B (<i>df</i> = 34)	32.49 (5.12)	32.45 (5.27)	29.43 (4.67)	33.58 (6.43)	33.62 (6.51)	31.88 (6.26)	34.10 (7.34)	34.47 (7.65)	33.82 (7.25)						
Type I error rate	0.00	0.01	0.00	0.02	0.03	0.01	0.04	0.04	0.03						
$RMSEA_B$	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)						
CFI_B	0.90 (0.20)	0.95 (0.12)	0.99 (0.05)	0.95 (0.08)	0.97 (0.06)	0.98 (0.04)	0.97 (0.04)	0.98 (0.03)	0.99 (0.02)						
TLL_B	0.88 (0.23)	0.93 (0.14)	0.98 (0.06)	0.93 (0.11)	0.96 (0.08)	0.98 (0.05)	0.96 (0.06)	0.97 (0.04)	0.98 (0.03)						
$SRMR_B$	0.15 (0.02)	0.12 (0.02)	0.10 (0.01)	0.11 (0.01)	0.09 (0.01)	0.07 (0.01)	0.07 (0.01)	0.06 (0.01)	0.05 (0.01)						
<i>W-l-s</i>															
χ^2_W (<i>df</i> = 34)	34.01 (6.54)	32.91 (6.28)	28.51 (5.12)	34.19 (7.61)	33.29 (7.08)	30.16 (6.24)	34.05 (7.69)	33.45 (7.62)	32.33 (7.14)						
Type I error rate	0.02	0.01	0.00	0.04	0.04	0.01	0.04	0.04	0.03						
$RMSEA_W$	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)						
CFI_W	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)						
TLL_W	0.98 (0.03)	0.99 (0.01)	1.00 (0.00)	0.99 (0.02)	1.00 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)						
$SRMR_W$	0.08 (0.01)	0.05 (0.01)	0.03 (0.00)	0.06 (0.01)	0.04 (0.01)	0.02 (0.00)	0.04 (0.01)	0.03 (0.00)	0.02 (0.00)						
M_B															
<i>B-l-s</i>															
χ^2_B (<i>df</i> = 35)	39.41 (7.17)	43.50 (7.89)	42.23 (8.13)	51.65 (10.64)	62.58 (13.08)	69.34 (14.90)	78.01 (16.90)	104.72 (21.25)	128.80 (24.49)						
Power	0.08	0.22	0.17	0.53	0.83	0.91	0.96	1.00	1.00						
$RMSEA_B$	0.04 (0.03)	0.06 (0.04)	0.06 (0.04)	0.06 (0.03)	0.09 (0.02)	0.10 (0.02)	0.08 (0.02)	0.10 (0.02)	0.11 (0.02)						
CFI_B	0.68 (0.29)	0.66 (0.25)	0.75 (0.21)	0.64 (0.20)	0.60 (0.17)	0.61 (0.16)	0.60 (0.15)	0.56 (0.12)	0.55 (0.11)						
TLL_B	0.61 (0.35)	0.57 (0.30)	0.68 (0.26)	0.53 (0.26)	0.49 (0.22)	0.50 (0.20)	0.49 (0.19)	0.44 (0.16)	0.42 (0.14)						
$SRMR_B$	0.18 (0.03)	0.15 (0.02)	0.14 (0.02)	0.14 (0.02)	0.13 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.01)	0.11 (0.01)						
M_W															
<i>W-l-s</i>															
χ^2_W (<i>df</i> = 35)	77.42 (18.01)	124.95 (27.24)	195.40 (37.17)	133.45 (27.98)	240.39 (42.16)	480.07 (64.48)	247.27 (40.44)	485.45 (60.09)	1112.66 (97.98)						
Power	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
$RMSEA_W$	0.05 (0.01)	0.05 (0.01)	0.04 (0.01)	0.05 (0.01)	0.05 (0.01)	0.05 (0.00)	0.05 (0.01)	0.06 (0.00)	0.06 (0.00)						
CFI_W	0.74 (0.10)	0.73 (0.07)	0.75 (0.04)	0.72 (0.07)	0.72 (0.05)	0.73 (0.03)	0.72 (0.05)	0.72 (0.03)	0.73 (0.02)						
TLL_W	0.66 (0.13)	0.65 (0.09)	0.68 (0.06)	0.64 (0.09)	0.65 (0.06)	0.66 (0.04)	0.64 (0.06)	0.64 (0.04)	0.65 (0.03)						
$SRMR_W$	0.13 (0.02)	0.12 (0.01)	0.11 (0.01)	0.12 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.00)						

Note. M_C = Correctly specified hypothesized model. M_B = Misspecification in between-level model only. M_W = Misspecification in within-level model only. B-l-s = Between-level specific. W-l-s = within-level specific. NG = Numbers of groups. GS = Group size. ICC = Intraclass correlation coefficient. The values of χ^2 statistics were reported to two decimal places for sake of space.

Acknowledgements This work was supported by the Ministry of Science and Technology, Taiwan (Grant Number MOST 110-2511-H-003-006-MY2) for John J. H. Lin.

Data availability Supplementary materials, including the simulation program and an example dataset are available at <https://github.com/johnjrhunglin/ML-SEM-categorical-fit-index>. Also, the annotated Mplus code for both correctly specified and misspecified models is available as well.

References

- Asparouhov, T., & Muthen, B. (2007). *Computationally efficient estimation of multilevel high-dimensional latent variable models*. Paper presented at the proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology.
- Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(2), 211–240. <https://doi.org/10.1080/10705510801922340>
- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34(2), 181–197. <https://doi.org/10.1207/S15327906MB340203>
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605–634. <https://doi.org/10.1146/annurev.psych.53.100901.135239>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Depaoli, S., & Clifton, J. P. (2015). A bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 327–351. <https://doi.org/10.1080/10705511.2014.937849>
- DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 425–438. <https://doi.org/10.1080/10705511.2014.915373>
- Finney, S. J., DiStefano, C., Hancock, G., & Mueller, R. (2006). Structural equation modeling: A second course. *LAP-Information Age Publishing Inc, United States, Chapter Non-normal and categorical data in structural equation modelling*, 269–314.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. <https://doi.org/10.1037/1082-989x.9.4.466>
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 625–641. <https://doi.org/10.1080/10705510903203573>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via monte carlo simulation. *Psychological Methods*, 21(1), 93–111. <https://doi.org/10.1037/met0000064>
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3), 319–336. <https://doi.org/10.1037/a0024917>
- Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(1), 1–27. <https://doi.org/10.1080/10705510802561279>
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Taylor & Francis.
- Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), 157–174. https://doi.org/10.1207/S15328007SEM0802_1
- Hsu, H.-Y. (2009). Testing the effectiveness of various commonly used fit indices for detecting misspecifications in multilevel structural equation models. (Unpublished doctoral dissertation). Texas A&M University.
- Hsu, H.-Y., Kwok, O., Acosta, S., & Lin, J.-H. (2015). Detecting misspecified multilevel SEMs using common fit indices: A monte carlo study. *Multivariate Behavioral Research*, 50(2), 197–215. <https://doi.org/10.1080/00273171.2014.977429>
- Hsu, H.-Y., Lin, J. J. H., Kwok, O., Acosta, S., & Willson, V. (2017). The impact of intra-class correlation on the effectiveness of level-specific fit indices in multilevel structural equation modeling (MSEM): A Monte Carlo study. *Educational and Psychological Measurement*, 77(1), 5–31. <https://doi.org/10.1177/0013164416642823>
- Hsu, H.-Y., Lin, J., Skidmore, S., & Kim, M. (2018). Evaluating fit indices in a multilevel latent growth curve model: A monte carlo study. *Behavior Research Methods*, 51. <https://doi.org/10.3758/s13428-018-1169-6>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 user's reference guide* (2nd ed.). Scientific Software International.
- 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://doi.org/10.1080/00273171.2016.1228042>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
- Lee, B., & Sohn, W. (2022). Testing the performance of level-specific fit evaluation in MCFA models with different factor structures across levels. *Educational and Psychological Measurement*, 82(6), 1153–1179. <https://doi.org/10.1177/00131644211066956>
- Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, 23(2), 318–336. <https://doi.org/10.1037/met0001122>
- Muthen, B. O. (1990). *Mean and covariance structure analysis of hierarchical data*. Paper presented at the psychometric society meeting, Princeton. UCLA statistics series 62 retrieved from <https://escholarship.org/uc/item/1vp6w4sr>
- Muthen, B. O. (1993). Goodness of fit with categorical and other non-normal variables. *SAGE Focus Editions*, 154, 205–205.
- Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22(3), 376–398.
- Muthen, L. K., & Muthen, B. O. (2017). *Mplus user's guide* (7th ed.). Muthén & Muthén.

- Muthen, L. K., & Muthen, B. O. (1998–2015). *Mplus user's guide* (7th ed.). Muthén & Muthén.
- Navruz, B. (2016). *The behaviors of robust weighted least squares estimation techniques for categorical/ordinal data in multilevel CFA models*. (Doctoral dissertation). Texas A&M University. Retrieved from <https://hdl.handle.net/1969.1/157931>
- Nestler, S. (2013). A monte carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 127–143. <https://doi.org/10.1111/j.2044-8317.2012.02044.x>
- Nye, C. D., & Drasgow, F. (2011). Assessing goodness of fit: Simple rules of thumb simply do not work. *Organizational Research Methods*, *14*(3), 548–570. <https://doi.org/10.1177/1094428110368562>
- Padgett, R. N., & Morgan, G. B. (2020). Multilevel CFA with ordered categorical data: A simulation study comparing fit indices across robust estimation methods. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*. <https://doi.org/10.1080/10705511.2020.1759426>
- 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>
- Rappaport, L. M., Amstader, A. B., & Neale, M. C. (2020). Model fit estimation for multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 318–329. <https://doi.org/10.1080/10705511.2019.1620109>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rigdon, E. E. (1996). Cfi versus rmsea: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(4), 369–379. <https://doi.org/10.1080/10705519609540052>
- Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods*, *43*(4), 1066–1074. <https://doi.org/10.3758/s13428-011-0115-7>
- Ryu, E. (2014). Model fit evaluation in multilevel structural equation models. *Frontiers in Psychology*, *5*, 1–9. <https://doi.org/10.3389/fpsyg.2014.00081>
- 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>
- Savalei, V. (2010). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a monte carlo comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 241–264. <https://doi.org/10.1080/10705511003659375>
- Savalei, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, *72*(6), 910–932. <https://doi.org/10.1177/0013164412452564>
- Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 201–223. <https://doi.org/10.1111/j.2044-8317.2012.02049.x>
- Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology*, *5*, 1–11. <https://doi.org/10.3389/fpsyg.2014.00181>
- Skrondal, A. (2000). Design and analysis of monte carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, *35*(2), 137–167. https://doi.org/10.1207/S15327906MBR3502_1
- Stapleton, L. M., McNeish, D. M., & Yang, J. S. (2016a). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist*, *51*(3–4), 317–330. <https://doi.org/10.1080/00461520.2016.1207178>
- Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016b). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. <https://doi.org/10.3102/107698616646200>
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. <https://doi.org/10.1007/BF02291170>
- Wu, J. Y., Lin, J. J. H., Nian, M. W., & Hsiao, Y. C. (2017). A Solution to modeling multilevel confirmatory factor analysis with data obtained from complex survey sampling to avoid conflated parameter estimates. *Frontiers in Psychology*, *8*, 19. <https://doi.org/10.3389/fpsyg.2017.01464>
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428. <https://doi.org/10.3758/s13428-018-1055-2>
- Yang-Wallentin, F., Joreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(3), 392–423. <https://doi.org/10.1080/10705511.2010.489003>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.