ABSTRACT

Title of Dissertation: EVALUATING MODEL FIT FOR LINEAR-LINEAR PIECEWISE MULTILEVEL LATENT GROWTH CURVE MODELS

Yuan Zhang, Doctor of Philosophy, 2015

Dissertation directed by: Dr. Hong Jiao, Department of Human Development and Quantitative Methodology

This dissertation examines the sensitivity of six fit indices in detecting various types of misspecifications in the application of a linear-linear piecewise multilevel latent growth curve model that uses continuous multivariate normal data. The study results show that all fit indices are more sensitive to misspecifications on the within level than those on the between level structure of the model. On the within level, all fit indices are more sensitive to the misspecification in the covariance structure than that in the residual structure; on the between level, all fit indices are more sensitive to the misspecification in the marginal mean structure than that in the covariance structure. Actually, none of the fit indices are practically significantly sensitive to the misspecification in the between-level covariance structure. Partially-saturated estimation method helps *NFI*, *TLI*, and *Mc* to be sensitive to the appropriate sample size when evaluating the misspecification in the between-level covariance structure;

however, it helps none of the fit indices when detecting models misspecified in the between-level covariance structure.

All fit indices are principally influenced by the severity of misfit if it happens on the within level; however, they are primarily affected by group size if the misspecification occurs at the between level. When severity level increases, all fit indices have more power to detect misspecification in the within-level covariance structure. When group size increases, *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are more likely to commit Type II errors in detecting misspecifications in the marginal mean structure and in both the marginal mean and the covariance structures. Compared with other fit indices, *NFI* is most vulnerable to sample size and least sensitive to severity level of misfit. *SRMR*, however, behaves differentially from all other fit indices in that it is most sensitive to the intraclass correlation coefficient when detecting studied misspecifications on the between level structure. Furthermore, the recommended cutoff values lead to high Type II errors for all fit indices in detecting various types of misspecifications, and it is infeasible to find a substitute new set of criteria based on the current data conditions.

EVALUATING MODEL FIT FOR LINEAR-LINEAR PIECEWISE MULTILEVEL LATENT GROWTH CURVE MODELS

by

Yuan Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Hong Jiao, Chair
Professor George Macready
Professor Laura Stapleton
Professor Jeffrey Patton
Professor Xin He
Professor Jing Lin

Acknowledgements

First of all, I would like to thank my advisor, Dr. Hong Jiao, for her immense care and patience with me during this arduous journey. Without her endless and extensive support, I would never have come to this topic and finished this dissertation.

I would also like to thank Dr. George Macready, who continuously inspired me and helped me on my academic study. Had it not been his guidance, I would not have passed the preliminary stage of my study in this completely new field. He also showed me a brilliant world of teaching and offered me so many opportunities in my personal growth, to which I highly appreciate.

Additionally, I would like to thank Dr. Robert Lissitz, who led me into the world of quantitative research. His insights and experience inspired me to explore interesting educational phenomena, during which I grasped comprehensive research methods and developed solid programming skills.

My sincere thanks also go to Dr. Laura Stapleton, Dr. Jeffrey Patton, and Dr. Xin He for their agreement to serve on my committee and provide constructive comments.

Lastly, I want to thank my family members for their support and encouragement to pursue my PhD degree.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Growth curve modeling has increased rapidly over the past 30 years in social and behavioral sciences (e.g., education, clinic, psychology, and sociology). Using the same or equated measures on individuals and/or groups repeatedly, growth curve modeling enables researchers to predict individuals and groups' changes over time, to investigate inter-group, inter-individual, and intra-individual variations in their development, and to examine the influence of background variables or treatment effects on the variation of individual and group growth trajectories (Bollen & Curran, 2006; Browne & du Toit, 1991; Bryk & Raudenbush, 1987; McArdle, 2009; Preacher, Wichman, MacCallum & Briggs, 2008; Raudenbush & Bryk, 2002; Singer & Willett, 2003).

Two major approaches arise in the course of the evolution of growth curve modeling. One is multilevel modeling (also known as hierarchical linear model or mixed-effects model) and the other is latent growth modeling (also known as latent growth curve model or latent trait model). Compared to multilevel modeling, latent growth modeling is more flexible in testing different research hypotheses[1] (e.g., Curran, 2000; Duncan, Duncan, Strycker, Li, & Alpert, 1999; McArdle & Bell, 2000; Raudenbush, 2001). In addition, it is capable of simultaneously modeling outcomes in multiple disciplines. Compared with multilevel confirmatory factor analyses modeling, growth curve modeling, including the multilevel latent growth curve

[1] Compared to multilevel modeling, limited time patterns are allowed in latent growth modeling.

models, include a mean structure into the model which enables researchers to track both individual and group level subjects change over time.

Previous researches on growth curve modeling often employ linear, polynomial, exponential, or logistic functional forms (e.g., Browne, 1993; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Markianidou & Weeren, 2011) that assume subjects change in a smooth and uninterrupted manner. This assumption, however, is often violated in real practice (e.g., in clinical and experimental practices). Hence, piecewise growth curve models were introduced to incorporate separate growth profiles for different developmental stages. Compared with other functional forms, piecewise growth curve models allow the growth slopes to relate differently to predictors, outcomes, or time-varying covariates, while presenting either the same, or a different, average growth shape (Dialloa & Morin, 2014).

Among piecewise growth curve models, linear piecewise models have gained much attention due to their simplicity in specification and interpretation (Flora, 2008; Kwok, Luo, & West, 2010). If a latent growth curve model adopts a linear-linear piecewise functional form, it becomes a linear-linear piecewise latent growth curve model. This model can be further complicated by incorporating multilevel latent structures to explain the growth trajectories associated with multilevel data (e.g., students nested within schools). In educational setting, this linear-linear piecewise multilevel latent growth curve modeling is especially useful in that it could be used to evaluate school and teacher effects. To be more specific, educators may expect students from different classes or schools to vary from each other; however, they would also hope that the ability of students from the same class or school reach homogeneity after a period of study. Compared with other models, this inconsistent change pattern and speed in student ability can be easily captured by linear-linear piecewise multilevel latent growth curve modeling.

Although linear-linear piecewise multilevel latent growth curve model is of great usage in educational and behavioral sciences (Muthen & Muthen, 1998-2012), methodological research is still absent concerning the evaluation of its proper application. An essential issue concerning the appropriate use of a statistical model is to evaluate the adequate fit of a hypothesized model. Within the structural equation modeling (SEM) framework, it is possible to judge the fit of a hypothesized model relative to a saturated baseline model by using the practical fit indices such as the Normed fit index (Bentler & Bonett, 1980), Tucker-Lewis Index (Tucker & Lewis, 1973), Comparative Fit Index (Bentler, 1989, 1990), McDonald's Centrality Index (McDonald, 1989), Root Mean Square Error of Approximation (Steiger & Lind, 1980; Steiger, 1989), and Standardized Root Mean Square Residual (Jöreskog & Sörbom, 1981; Bentler, 1995).

A few studies have been conducted to evaluate the performance of the practical fit indices in detecting model misspecifications concerning single-level confirmatory factor analytical models, multilevel confirmatory factor analytical models, or latent growth curve models. Their results suggest that those indices do not perform equally well across different types of misspecification and their accuracy to detect modeling misfit may be influenced by such factors as the degree of misfit and sample size. Another issue involving the use of the fit indices is the appropriateness of the suggested cutoff values, whose generalizability is also limited to the statistical model and conditions included in a particular study (Hu & Bentler, 1995). Hence the purpose of this dissertation is to investigate the sensitivity of six commonly used practical fit indices (i.e., *NFI*, *TLI*, *CFI*, *Mc*, *RMSEA*, and *SRMR*) in detecting model misspecifications in a linear-linear piecewise multilevel latent growth curve model which is based on continuous normally distributed outcomes.

The dissertation is organized as follows. In Chapter II, the rationale of linear-linear piecewise multilevel latent growth curve models with normally distributed outcome is first explained, followed by the summary of factors that might influence the estimation and evaluation of this type of model. Subsequently, the characteristics of the six practical fit indices as well as previous research on their performance are reviewed. Chapter III introduces the research design of the study, followed by the real data analysis of the Longitudinal Study of American Youth (LSAY), which provides population values for the simulation study. Chapter IV explains the simulation results and the final chapter makes a conclusion about the performance of the fit indices concerning the evaluation of linear-linear piecewise multilevel latent growth curve models.

Chapter 2: Literature Review

2.1 Linear-Linear Piecewise Multilevel Latent Growth Curve Models

2.1.1 Arise of piecewise multilevel latent growth curve models

To understand reliably and validly the causes, development, and consequences of human behavior has long been the primary goal for behavioral sciences. Before the introduction to growth curve modeling, researchers and practitioners have used various methods to evaluate people's growth over time, such as the repeated measures *t* tests, the analysis of variance (ANOVA), the analysis of covariance (ANCOVA), the multivariate analysis of variance (MANOVA), the multivariate analysis of covariance (MANCOVA), and multiple regression (Curran & Hussong, 2003; Hedeker & Gibbons, 2006). A commonality among all these models is that they tend to be considered as fixed-effects models. In other words, systematic relations are evaluated by averaging across individuals and the only source of random variation lies in the residuals (Curran & Hussong, 2003). In addition, ANOVA is highly constrained by its strict assumptions. To be more specific, it assumes the variance-covariance matrix to meet the requirement of compound symmetry, or in real practice its sufficient condition – compound symmetry, which implies that the variances of measurements at each time period are equal. This is highly unrealistic, if not at all impossible for many longitudinal studies (Kwok, West, & Green, 2007).

In order to solve the above problems, growth curve modeling has been proposed to simultaneously estimate the intra-individual growth trajectories and inter-individual differences in those growth parameters influenced by either time-varying or time-invariant covariates. Two major approaches, namely, hierarchical linear modeling (also known as mixed-effects growth modeling or multilevel growth modeling, see Bryk & Raudenbush, 1987; Raudenbush & Bryk, 2002; Singer & Willett, 2003) and structural equation modeling (also known as latent growth curve modeling, see Bollen & Curran, 2006; Duncan, Duncan, & Strycker, 2006; McArdle, 1988, 2009; McArdle & Epstein, 1987; Meredith & Tisak, 1990; Preacher et al., 2008) emerge consequently, which share certain similarities but are also characterized by distinct differences (Bauer, 2003; Curran, 2003; Raudenbush, 2001; Willett & Sayer, 1994). Briefly speaking, the multilevel growth curve modeling approach treats repeated measures as nested within each individual and decomposes the explained variances in a univariate outcome into its within-group and between-group components. However, the latent growth curve modeling approach incorporates repeated measures as multiple indicators on one or more latent factors to characterize the unobserved growth trajectories and explain the effects of sets of predictors from different levels (Heck & Thomas, 2008; Hsu, 2009). In many situations, the multilevel and the latent growth curve modeling approaches produce numerically identical results (Curran, Obeidat, & Losardo, 2010); however, the latter approach shows greater precision and more flexibility in modeling structural relationships between variables in that 1) it provides the option to exclude measurement error from both the predictors and the outcomes (Muthén, 2000; Palardy, 2003), 2) it incorporates multivariate latent and measured variables (Heck & Thomas, 2008; Muthén, 2000; Palardy, 2003), and 3) it allows multiple indicators for a single outcome or multiple outcomes at each time point (Kaplan & Georege, 1998; Muthén, 2000; Palardy, 2003; Sayer & Cumsille, 2001). Further, it enables simultaneous evaluation of direct, indirect[2], and total effects among outcome and predictor variables (Kline, 2011; Muthén, 2000)[3].

---

[2] HLM can estimate indirect effects between randomly varying covariates, but it cannot estimate covariance relationships between fixed effects at any level of analysis (Palardy, 2003).

[3] In multilevel growth curve models, time is introduced as a fixed predictor; however, it is introduced as factor loadings in multilevel latent growth curve models. Essentially, this makes the multilevel growth curve modeling a univariate approach, with time points being treated as observations of the same variable, whereas the multilevel latent growth curve modeling a multivariate approach, with time points being treated as separate variables (Hox & Stoel, 2005; Stoel & van den Wittenboer, 2003). This configuration of the time variable has certain consequences for the growth analysis that can be applied to different types of data. In other words, it is more convenient to use multilevel latent growth curve modeling with time-structured panel data – data which are complete and which measure subjects under an identical time scheme. In addition, since SEM models use sample-level data (i.e., sample means and covariances) instead of individual-level data with its parameter estimation, latent growth curve modeling assumes that the means and variances for the observed measures across measurement occasions are conditional on time and the covariance among the observed measures at any two measurement occasions is conditional on time. A single pooled-sample covariance matrix fails to capture such time dependence if individuals vary considerably at the first measurement occasion or if the time interval across measurement occasions is not fixed across individuals. Under such circumstances, the latent growth curve models would provide inaccurate estimates of the random components, including the variance estimate of the intercept factor and the covariance between the slope and intercept factors (Mehta & West, 2000).

To date, the majority of applications of growth modeling have used linear models (e.g., Bryk & Raudenbush, 1987; Duncan, Duncan, Alpert, Hops, Stoolmiller, & Muthén, 2010 ; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; Muthén, 1991, 1997a, 1997b; Muthén & Asparouhov, 2008, 2011; Muthén & Khoo, 1998). These models are parsimonious, simple to interpret, and often provide an adequate representation of the growth process when there are a small number of measurement waves (i.e., 3 or 4 repeated measures) (Kwok, Luo, & West, 2010). Longitudinal data, however, frequently show nonlinear patterns of change over time, especially when there is a relatively large number of measurement occasions (i.e., five or more repeated measures). Hence, nonlinear functions such as polynomial, logistic, exponential, or Gompertz are used to capture the nonlinear components of a change (e.g., Browne, 1993; Huttenlocher et al., 2011). These functional forms, however, are typically more complicated to estimate and challenging to interpret, especially when covariates are introduced into the model to explain individual heterogeneity in the growth factors (Flora, 2008; Kwok, Luo, & West, 2010). Alternatively, piecewise linear models represent a parsimonious, easily implemented approach to describing nonlinear trajectories. In addition to its straightforward interpretation of the additional slope factor, they can be explicitly tailored to describe changes during a specific window of time over the full time span, or they can target particular changes after a theoretically important time point (Flora, 2008). Various numbers of knots and different functional forms, especially the linear functions within each time segment may be used to accurately reflect subjects’ growth trajectories.

In addition to turning to growth curve modeling, researchers have also noticed the multilevel nature of the behavioral and social science data (e.g., students are nested within classroom, classrooms are nested within schools, and schools are further nested within school districts), which may result from the application of complex sampling procedures or from their practice to administer interventions to higher level units in experimental studies (Ryu, 2008). Statistically, that nestedness implies that individuals within a group respond more similarly than those from another group, or the residuals are correlated to each other. This violation of the independence assumption, if ignored, often results in underestimated standard errors, inflated chi-square values and Type I error rate, and biased parameter estimates (e.g., underestimated group-level varaince) (Hox, 1998; Muthén & Satorra, 1995). This is particularly pronounced when the intraclass correlation is large and the number of individuals within each group (i.e., cluster size) is large (Muthén, 1997a). In addition to this statistical assumption violation, the within-level structure and the between-level structure of a multilevel model may adopt different functional forms and present different practical meanings. For example, the performance of students nested within the same school may either grow or decrease over a period of time. However, the performance of the school may keep constant over the same period of time, suggesting that the school does not improve or regress over that time.

Historically, two methods were adopted to deal with multilevel data structure (Muthén & Satorra, 1995). One is the aggregation method (i.e., ignoring information at the individual level by averaging measurement scores within groups, which implies that the within-group variability of the aggregated variable is assumed to be zero

(Barr, 2008)) and the other is the disaggregation method (i.e., ignoring the hierarchical structure altogether and using the measurement scores at the individual-level only). Both methods, however, may lead to statistical fallacies and produce inaccurate parameter estimates and research conclusions. To be more specific, using the aggregation method may commit ecological fallacy (i.e., an inference about the nature of individuals are deduced from group-level information (Robinson, 1950)) and produce underestimated statistical power, incorrect sample size and weighting of groups, and unreliable group-level information (Alker, 1969; Blalock, 1979; Chou, Bentler, & Pentz, 2000; Diez-Roux, 1998; Firebaugh, 1978; Hofmann, 1997, 2002; Klein & Kozlowski, 2000; Krull & MacKinnon, 2001; Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov, & Muthen, 2008; Preacher, Zyphur, & Zhang, 2010; van de Vijver & Poortinga, 2002). Using the disaggregation method confounds the within- and between-group relationships (Cronbach, 1976), which may result in atomistic fallacy (i.e., an inference about causal relationships in groups are made on the basis of relationships observed in individuals (Diez-Roux, 1998)) and yield underestimated standard errors, inflated chi-square values, Type I error rate and power for indirect effects, and underestimated parameter estimates (Chou, Bentler, & Pentz, 2000; Hox, 1998; Julian, 2001; Krull & MacKinnon, 1999; Muthén & Satorra, 1995; Ryu, 2008; Preacher, Zyphur, & Zhang, 2010). In addition, Simpson's paradox (i.e., the direction of the relationship reverses) may occur when collapsing groups from heterogeneous populations as if they were from a homogeneous population (Simpson, 1951). To avoid those fallacies, multilevel modeling is proposed, which partitions the outcome variable variance into individual and group level components respectively, with the aim to promote the correct conceptualization of the influence of different sets of predictors at different levels and to clarify the degree to which the outcome variance is due to differences between individuals as opposed to differences between groups (Palardy, 2003; Raudenbush & Bryk, 2002).

In addition to the evolution in theories mentioned above, both general (e.g., STATA) and specialized (e.g., Mplus, EQS) estimation procedures in software programs have been developed to estimate model parameters of piecewise multilevel latent growth curve model with less constraints (e.g., time coding). All these factors, together with the inherent modeling flexibility and precision, leads to an attention to linear-linear piecewise multilevel latent growth curve modeling in methodological and practical researches.

2.1.2 Linear-linear piecewise multilevel latent growth curve models

A linear-linear piecewise multilevel latent growth curve model is a simple extension of the linear multilevel latent growth curve model. By introducing an additional slope factor, it can be used to describe two pieces of linear change occurring over two separate time segments. The matrix form of a linear-linear piecewise multilevel latent growth curve model is the same as that of a linear multilevel latent growth curve model. Following Muthén’s notation, an unconditional linear-linear piecewise multilevel latent growth curve model is expressed as:

$$\mathbf{Y_{gi}} = \mathbf{v} + \mathbf{\Lambda_g} \boldsymbol{\eta}_{\mathbf{gi}} + \boldsymbol{\varepsilon}_{\mathbf{gi}} \tag{1}$$

$$\boldsymbol{\eta}_{gi} = \boldsymbol{\alpha}_g + \mathbf{B}_g \boldsymbol{\eta}_{gi} + \boldsymbol{\zeta}_{gi} \tag{2}$$

$$\boldsymbol{\eta}_g = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_g + \boldsymbol{\zeta}_g \tag{3}$$

with $\boldsymbol{\varepsilon}_{gi} \sim MNV\left(\mathbf{0},\boldsymbol{\Theta}\right), \boldsymbol{\zeta}_{gi} \sim MNV\left(\mathbf{0},\boldsymbol{\Psi}_{W}\right), \boldsymbol{\zeta}_{g} \sim MVN\left(\mathbf{0},\boldsymbol{\Psi}_{B}\right)$[4]

$$\operatorname{cov}\left(\boldsymbol{\varepsilon}_{gi},\boldsymbol{\zeta}_{gi}\right)=\mathbf{0}, \operatorname{cov}\left(\boldsymbol{\varepsilon}_{gi},\boldsymbol{\zeta}_{g}\right)=\mathbf{0}, \operatorname{cov}\left(\boldsymbol{\zeta}_{g},\boldsymbol{\zeta}_{gi}\right)=\mathbf{0}$$

Usually, the measurement time points are constrained to be equal across groups (i.e., $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_{g}$), making the factor loadings to be a (T×3) matrix. This practice makes it possible to convert the time variable into factor loadings (Muthén, 1997a, 1997b; Muthén & Khoo, 1998), which suggests that the same measures are used across groups and the same measurement variables are important in capturing the cross-group variability (Muthén, Khoo, & Gustafsson, 1997). In addition, by constraining the within cluster means as well as the between cluster residuals to be zero, Equations (1) – (3) imply that[5],

$$\boldsymbol{\mu}\left(\boldsymbol{\theta}\right)=\mathbf{v}+\boldsymbol{\Lambda}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\alpha} \tag{4}$$

$$\begin{aligned}\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)&=\boldsymbol{\Lambda}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Psi}_{\mathbf{B}}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Lambda}'+\boldsymbol{\Theta}_{\mathbf{B}}+\boldsymbol{\Lambda}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Psi}_{W}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Lambda}'+\boldsymbol{\Theta}_{W}\\&=\boldsymbol{\Lambda}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Psi}_{\mathbf{B}}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Lambda}'+\boldsymbol{\Lambda}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Psi}_{W}\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Lambda}'+\boldsymbol{\Theta}\end{aligned} \tag{5}$$

The specification that the mean appears only on the between-level for multilevel latent growth curve models is in line with the conventional single-level analysis in that means are specified for the level of variation containing independent observations. Actually, it is generally impossible to identify the within-level factor mean and unnecessary to let it deviate from zero because such across-population differences can be captured in the between-level factors (Muthén, Khoo, &

[4] According to Muthén & Asparouhov (2008), the distributional assumption for ε_{gi} is determined by the type of observed variable included in the model.

[5] The inclusion of a mean structure in growth curve modeling is the major difference between a multilevel latent growth curve model and a multilevel confirmatory factor analysis model.

Gustafsson, 1997). The specification of constraining the between-level residuals to be zero is to make the model in line with the conventional multilevel growth curve model, which however practically suggests that "the conventional model [i.e., multilevel growth curve models] tries to absorb the residual variances [on the between-level] into the slope growth factor variance" (Muthén and Asparouhov, 2011, pp. 31).

One of the key features of a piecewise model is the decision of the transition point, which is also known as the knot. Its value can be either fixed or estimated (e.g., Flora, 2008; Harring, Cudeck, & du Toit, 2006; Kwok, Luo, & West, 2010; Kohli & Harring, 2013). For a linear-linear piecewise multilevel latent growth curve model with 6 equally spaced time points where the location of the knot occurs at the 4th time point, its relationship can be illustrated by a traditional path diagram (see Figure 1).

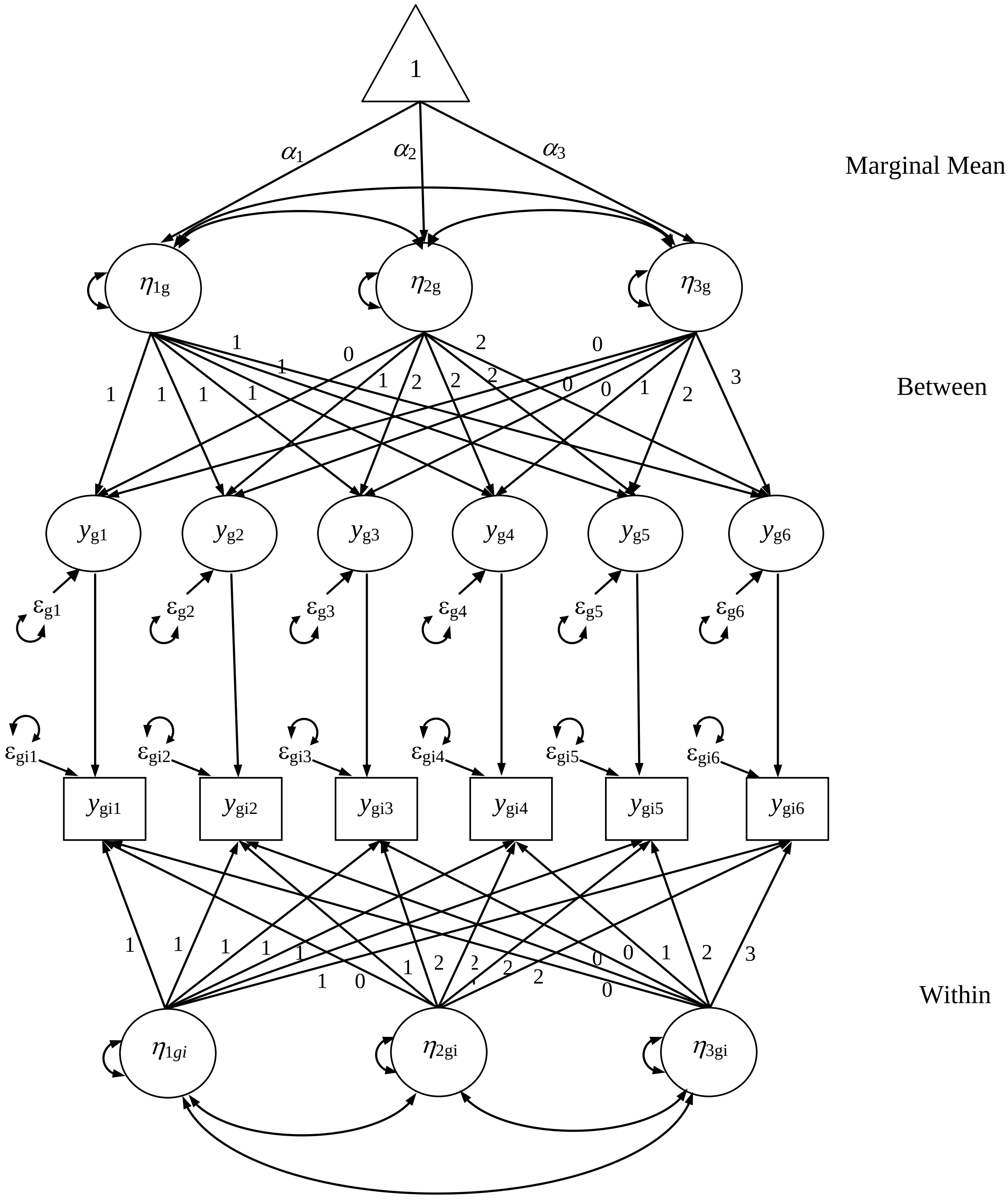


Figure 1: Path diagram for a linear-linear piecewise multilevel latent growth curve model with fixed knot.

In this diagram, the factor loading matrix $\mathbf{\Lambda}_g$ is:

$$\mathbf{\Lambda}_g = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

meaning that all the subjects within groups share the same transition point. This is often seen in an educational system where all the subjects move from primary schools to high schools at almost the same age, or is observed in experimental studies where a treatment is given to all the subjects at a particular time point. This constraint, however, could be released by introducing individually varying knot locations (see Preacher & Hancock, 2010). Furthermore, the specification of the factor loadings can be changed by choosing a different reference point (e.g., instead of choosing the origin as the reference point, one may choose the knot position as the reference point) and time coding method (e.g., additive versus piecewise) (see Flora (2008) for further clarification).

2.1.3 Estimation method for linear-linear piecewise multilevel latent growth curve models

Following Muthén's notation (1989, 1990, 1997a), the maximum likelihood estimation of a linear-linear piecewise multilevel latent growth curve model can be briefly described as follows. Assume the independent outcomes as functions of the group- level (e.g., *z*) and individual-level (e.g., *y*) variables respectively, arrange the data vector as

$$\mathbf{d}_g^{'} = \left(z_g^{'}, y_{g1}^{'}, y_{g2}^{'}, ..., y_{gN_g}^{'} \right) \quad (6)$$

where g = 1, 2, ..., G, indicating independently observed groups and

i = 1, 2, …, N_g, indicating individual observations within group g

the mean and the covariance matrices are

$$\mu_{\mathbf{d_g}}^{'} = \left(\mu_{\mathbf{z}}^{'}, \mathbf{1}_{N_g}^{'} \otimes \mu_{\mathbf{y}}^{'} \right) \quad (7)$$

$$\Sigma_{\mathbf{d_g}} = \begin{pmatrix} \Sigma_{\mathbf{zz}} & symmetric \\ \mathbf{1}_{N_g} \otimes \Sigma_{\mathbf{yz}} & \mathbf{I}_{N_g} \otimes \Sigma_w + \mathbf{1}_{N_g} \mathbf{1}_{N_g}^{'} \otimes \Sigma_B \end{pmatrix} \quad (8)$$

where $\otimes$ denotes a Kronecker product

$\mathbf{I}_{\mathrm{Ng}}$ denotes an identity matrix of dimension ($N_g \times N_g$)

$\mathbf{1}_{\mathrm{Ng}}$ denotes a vector of N_g unit elements

Σ_W is the ($p \times p$) within-group covariance matrix for the y variables

Σ_B is the ($p \times p$) between-group covariance matrix for the y variables

Assuming multivariate normality of $\mathbf{d}_g$[6], the ML estimator minimizes the function

[6] Standard ML estimation assumes the endogenous variables to be multivariate normal, which implies that (1) all univariate distributions should be normal, (2) all bivariate scatterplots are linear, and (3) the distribution of residuals is homoscedastic. The normality assumption in ML estimation is critical in that if severely violated, the standard errors for parameter estimates tend to be too low and the model chi-square value tends to be either too high or too low, which may result in inflated Type I and Type II error rates (Bentler & Yuan, 1999; Kline, 2012). In addition, with categorical endogenous variables, the ML estimator tends to produce deflated parameter estimates and their standard errors (DiStefano, 2002). In that case, weighted least squares estimation (see Asparouhov & Muthén, 2007) is often suggested to efficiently provide correct model chi-square value, parameter estimates, and their standard errors (Kline, 2012).

$$F = \sum_{g=1}^{G} \left\{ \log \left| \Sigma_{\mathbf{d_g}} \right| + \left(\mathbf{d_g} - \mu_{\mathbf{d_g}} \right)' \Sigma_{\mathbf{d_g}}^{-1} \left(\mathbf{d_g} - \mu_{\mathbf{d_g}} \right) \right\} \tag{9}$$

which may be simplified as

$$\begin{aligned} F = & \sum_{d=1}^{D} G_d \left\{ \ln \left| \Sigma_{B_d} \right| + tr \left[\Sigma_{B_d}^{-1} \left(S_{B_d} + N_d \left(\bar{\nu}_d - \mu \right) \left(\bar{\nu}_d - \mu \right)' \right) \right] \right\} \\ & + \left(N - G \right) \left\{ \ln \left| \Sigma_W \right| + tr \left[\Sigma_W^{-1} S_{PW} \right] \right\} \end{aligned} \tag{10}$$

with
$$\Sigma_{B_d} = \begin{pmatrix} N_d \Sigma_{zz} & symmetric \\ N_d \Sigma_{yz} & \Sigma_W + N_d \Sigma_B \end{pmatrix}$$

$$S_{B_d} = N_d G_d^{-1} \sum_{k=1}^{G_d} \begin{pmatrix} z_{dk} - \bar{z}_d \\ \bar{y}_{dk} - \bar{y}_d \end{pmatrix} \left[\left(z_{dk} - \bar{z}_d \right)' \left(\bar{y}_{dk} - \bar{y}_d \right)' \right]$$

$$\bar{\upsilon}_d - \mu = \begin{pmatrix} \bar{z}_d - \mu_z \\ \bar{y}_d - \mu_y \end{pmatrix}$$

$$S_{PW} = \left(N - G \right)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{Ng} \left(y_{gi} - \bar{y}_g \right) \left(y_{gi} - \bar{y}_g \right)'$$

where *D* indicates the number of groups of a distinct size

d indicates a distinct group size category with group size N_{d}

G_{d} indicates the number of groups of that size

S_{Bd} indicates the between-group sample covariance matrix and

S_{PW} indicates the pooled-within sample covariance matrix

For the above estimation method, a multiple-group analysis is carried out for (*D*+1) groups, with the first *D* groups having sample size G_{d} and the last group having sample size (*N*-*G*) (Muthén, 1990, 1997a, 1997b). In other words, “equality constraints are imposed across the groups for the elements of the parameter arrays μ,

Σ_{zz}, Σ_{yz}, Σ_B, and Σ_W" (Muthén, 1997a, pp. 155). When this data balance assumption is violated, the ML estimation implies specifying a separate between-group model for each distinct group size (i.e., create a different 'group' for each set of groups with the same group size). These between-group models have different scaling parameters and different mean structures, and require equality constraints across all other parameters within that group (Muthén, 1994). Since conventional structural equation modeling software inverts matrices at each iteration with its estimation procedure (see Liang & Bentler, 2004; McDonald & Goldstein, 1989), the ML estimation becomes computationally heavy or even impossible when the number of distinct group size increases (e.g., the between-group covariance matrices are not positive definite because the number of variables is greater than the number of observations) (Duncan, Duncan, Alpert, Hops, Stoolmiller, and Muthén, 1997; Hox, 2010). Therefore, Muthén (1991, 1994) proposed an ad hoc estimator (i.e., MUML estimator) which ignored the variation in group size to increase the operationality of the ML estimator in real practice. By constraining $D = 1$, $G_d = G$, and $N_d = c$, the MUML estimator is defined as

$$F^{'} = G\left\{\ln\left|\Sigma_{B_c}\right| + tr\left[\Sigma_{B_c}^{-1}\left(S_B + c\left(\bar{v}-\mu\right)\left(\bar{v}-\mu\right)^{'}\right)\right]\right\} + \left(N-G\right)\left\{\ln\left|\Sigma_W\right| + tr\left[\Sigma_W^{-1}S_{PW}\right]\right\} \tag{11}$$

where c reflects the average group size,

$$c = \left[N^2 - \sum_{g=1}^{G} N_g^2\right]\left[N\left(G-1\right)\right]^{-1} \tag{12}$$

When data are balanced, the MUML estimator equals the ML estimator; when data are unbalanced, Muthén and his colleagues (Muthén, 1989, 1990, 1994, 1997a, 1997b; Muthén, Khoo, & Gustafsson, 1997) claimed that the ad hoc estimator produced parameter estimates, their standard errors and model chi-square value quite close to those obtained by the true ML estimator. McDonald (1994) also checked the equivalence of the ad hoc estimator to the ML estimator for unbalanced data, whose results indicated that both estimators led to the same conclusion for model inference. In contrast to those ideal conclusions, Hox and his colleagues conducted several simulation studies (Hox, 1993; Hox and Maas, 2001; Hox, Maas & Brinkhuis, 2010) to compare the performance of the MUML estimator to that of the ML estimator for unbalanced data. Their results suggested that the within-group part of the model posed no problem and the factor loadings were generally accurate, but the residual variances were underestimated and the standard errors were generally too small, leading to an inflated Type I error rate. In addition to that empirical work, Yuan and Hayashi (2005) showed analytically that the MUML standard errors and chi-square tests led to correct inferences only when the between-group sample size went to infinity and the coefficient of variation of the group sizes went to zero. With severely unbalanced data, the ad hoc estimator produced biased standard errors and significance tests, and this bias would not diminish when sample size increased.

2.2 SEM-Based Fit Indices

A major drawback associated with the chi-square test statistic is that it is sensitive to sample size (Fan, Thompson, & Wang, 1999; Hox, 2000; Sun, 2005; Widaman & Thompson, 2003): the power of the test increases with the increase in sample size. As a result, model fit assessment becomes very strict when sample size is large — a minimal discrepancy between the reproduced and the sample covariance matrices may become statistically significant. As stated by Bentler and Bonett (1980), the null hypothesis that $\Sigma(\theta) = \Sigma$ will always be rejected if sample size is large enough. In contrast, when the sample size is small, the test may fail to detect significant differences between the sample and the reproduced covariance matrices. The second drawback of the chi-square test statistic is that it favors complex models, meaning that its value always decreases when more parameters are freed for estimation until the model becomes saturated (Sun, 2005). Besides, chi-square test statistic is affected by the distribution of the data such that it increases with the increment of skewness and kurtosis of the variable. Consequently, a variety of alternative model fit indices have been developed, with the goal to produce measures that do not depend on sample size and address the distributional misspecification of the variable[7] (Bentler & Bonett, 1980). Therefore, two issues arise concerning the proper applications of the fit indices in model selection and evaluation. One is the determination of the adequacy of the fit indices under various data and model conditions (e.g., model misspecification, sample size, estimation method, violation of normality and independence assumptions, and model complexity); and the other is the selection of the "rules of thumb" cutoff criteria for the fit indices (Hu & Bentler, 1999). Nowadays, most researchers rely mainly on fit indices and their cutoff values suggested by Hu and Bentler (1999) as guidelines to justify the adequacy of hypothesized models (Hsu, 2009).

[7] Most goodness-of-fit indices still depend on sample size and also the distribution of the data, but the dependency is much smaller than that of the routine chi-square test (Hox, 2000).

2.2.1 Classification of the practical fit indices

Multiple methods have been proposed to classify the practical fit indices (i.e., the fit indices produced in contrast to the chi-square statistic, see Widaman & Thompson, 2003) into several categories (see Table 1). For example, Yuan (2005) classified the structural equation modeling fit indices into two major categories based on their construction procedure. That is, whether a fit index is defined directly through a likelihood ratio test or through the residuals in the mean and the covariance matrices. He argued that "all of the fit indices can be treated as weighted functions of residuals, but the fit indices that are defined through test statistics utilize theoretically more optimal weight functions" (Wu, West, & Taylor, 2009, pp. 190). In addition to this simple classification, a more systematic categorization was proposed by Tanaka (1993) who identified six dichotomous dimensions to address the multifaceted nature of the fit indices: (1) population-based versus sample-based, (2) simplicity versus complexity, (3) normed versus nonnormed, (4) absolute versus relative, (5) estimation method free versus estimation method specific, and (6) sample size independent versus sample size dependent. However, these dimensions are not totally independent from each other (Sun, 2005), hence a hierarchy of three different levels based on Tanaka's conception was proposed: discrepancy assumption level (i.e., whether or not a fit index is based on the assumption that $\Sigma(\theta) = \Sigma$, which divides the fit indices into sample-based and population based), model involvement level (i.e., whether or not a fit index involves another baseline model, which divides the fit indices into absolute and relative), and complexity adjustment level (i.e., whether or not a fit index is adjusted for model complexity, which divides the fit indices into adjusted and unadjusted). Of these levels, the discrepancy assumption level is the most fundamental, whereas the complexity adjustment level is the least fundamental. As is claimed by the author, “this classification scheme may facilitate the interpretation, comparison, and selection of fit indices” (Sun, 2005, pp. 243).

Table 1: Classifications and Definitions of the Fit Indices

Type (Hu & Bentler, 1995, 1998)	**Hierarchy** (Sun, 2005)	**Construction** (Yuan, 2005)	**Algebraic Definition**	**References**
Incremental Type 1	Sample-relative-unadjusted	Chi-square based	$NFI = \frac{T_b - T_h}{T_b}$	Bentler & Bonett (1980)
Incremental Type 2	Sample-relative-adjusted	Chi-square based	$TLI = \frac{T_b / df_b - T_h / df_h}{T_b / df_b - 1}$	Tucker & Lewis (1973) Benter & Bonett (1980)
Incremental Type 3	Population-relative-unadjusted	Chi-square based	$CFI = 1 - \frac{\max\left[\left(T_h - df_h\right), 0\right]}{\max\left[\left(T_h - df_h\right), \left(T_b - df_b\right), 0\right]}$	Bentler (1989, 1990)
Absolute	Population-absolute-unadjusted	Chi-square based	$M_c = \exp\left[-\frac{1}{2\left(\frac{T_h - df_h}{N-1}\right)}\right]$	McDonald (1989)
Absolute	Population-absolute-unadjusted	Chi-square based	$RMSEA = \sqrt{\max\left[\frac{T_h - df_h}{df_h\left(N-1\right)}, 0\right]}$	Steiger & Lind (1980) Steiger (1989)
Absolute	Sample-absolute-unadjusted	Residual-based	$SRMR = \sqrt{\frac{2\sum_{i=1}^{p}\sum_{j=1}^{i}\left(\frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ii}s_{jj}}\right)^2}{p\left(p+1\right)}}$	Jöreskog & Sörbom (1981) Bentler (1995)

Notes: 1) $T_b = \chi^2$ test statistic for the baseline model; $T_h = \chi^2$ test statistic for the hypothesized model; df_b = degrees of freedom for the baseline model; df_h = degrees of freedom for the hypothesized model; S_{ij} = observed covariances; $\hat{\sigma}_{ij}$ = reproduced covariances; S_{ii} and S_{jj} = observed standard deviations; p = number of variables; and N = sample size; 2) *NFI* = normed fit index; *TLI* = Tucker-Lewis index; *CFI* = comparative fit index; *Mc* = McDonald's centrality index; *RMSEA* = root-mean-square error of approximation; and *SRMR* = standardized root mean squared residual.

In addition to those grouping methods, a widely cited classification method was suggested by Hu and Bentler (1995, 1998), who divided the practical fit indices into two general classes: absolute and incremental fit indices. An absolute-fit index directly assesses how well an a priori model reproduces the sample data whereas an incremental fit index measures the proportionate improvement in fit by comparing a target model with a more restricted, nested baseline model. Regardless of using a reference model or not, both types of fit indices make an implicit or explicit comparison to a saturated model that exactly reproduces the observed covariance matrix (Hu & Bentler, 1999; Widaman & Thompson, 2003).

The incremental fit indices are further divided into three groups, namely, Types 1-3 (Hu & Bentler, 1998, pp. 426 – 427):

> "A Type 1 index uses information only from the optimized statistic *T*, used in fitting baseline (T_B) and target (T_T) models. *T* is not necessarily assumed to follow any particular distributional form, though it is assumed that the fit function *F* is the same for both models. A general form of such indices can be written as [Type 1 incremental indices = $\left|T_B - T_T\right| / T_B$].
>
> Type 2 and Type 3 indices are based on an assumed distribution of variables and other standard regularity conditions. A Type 2 index additionally uses information from the expected values of T_T under the central chi-square distribution. It assumes that the chi-square estimator of a valid target model follows an asymptotic chi-square distribution with a mean of df_T, where df_T is the degrees of freedom for a target model. Hence, the baseline fit T_B is compared with df_T, and the denominator in the Type 1 index is replaced by $\left(T_B - df_T\right)$. Thus, a general form of such indices can be written as [Type 2 incremental fit index = $\left|T_B - T_T\right| / \left(T_B - df_T\right)$].
>
> A Type 3 index uses Type 1 information but additionally uses information from the expected values of T_T or T_B, or both, under the relevant noncentral chi-square distribution.

When the assumed distributions are correct, Type 2 and Type 3 indices should perform better than Type 1 indices because more information is being used."

When comparing the performance of absolute and incremental fit indices, many absolute fit indices are relatively poor indicators of practical fit, as they are related too strongly to sample size. The root-mean-square error of approximation (Browne & Cudeck, 1993) and the centrality index (McDonald, 1989) represent two notable exceptions to this trend. In contrast, most of the commonly used incremental fit indices exhibit relative independence from sample size and thus are useful indices of practical fit (Widaman & Thompson, 2003). A scrutinization of the incremental fit indices reveals that there is a positive association between sample size and Type 1 incremental fit indices. To be more specific, Type 1 incremental fit indices tends to underestimate their asymptotic values and over-reject true models at small sample sizes. On the other hand, the Type 2 and Type 3 indices seem to be substantially less biased. The Type 2 and Type 3 incremental fit indices, in general, perform better than either the absolute or Type 1 incremental indices. Moreover, all the fit indices behave more consistently across estimation methods under the true-population model than under the misspecified models (Hu & Bentler, 1998).

2.2.2 Baseline model for relative fit indices

Specifying a correct baseline model is critical for relative fit indices in that a relative fit index with an incorrectly specified baseline model have no valid interpretation and may lead to biased inferences. Although several options exist for specifying a baseline model, an acceptable baseline model must satisfy the following conditions: (1) it must be nested within the hypothesized model, (2) it must estimate as few parameters as reasonable for the data, and (3) it must reproduce a nonzero variance and mean (if included in the analysis) for each manifest variable (Widaman & Thompson, 2003).

In view of latent growth curve modeling, the standard baseline model defined by *Mplus* (Muthén & Muthén, 1998-2012) is an independence model in which the covariances among the manifest variables are set to zero, but means and variances are unrestricted (Bentler & Bonett, 1980; Leite & Stapleton, 2011; Widaman & Thompson, 2003). Unfortunately, this standard baseline model is inappropriate because it freely estimates the outcomes' means and therefore is not nested within the linear latent growth curve model. Instead, an appropriate baseline model should state that there is no growth.

Widaman and Thompson (2003) specified two acceptable baseline models that may be used for most of the commonly used growth curve models (i.e., all of the polynomial models and linear piecewise models). Both null models are based on an intercept-only growth model, with the latter being more restricted than the former one. The former model constrains the means of all the manifest variables to be equal and only freely estimates the residual variance for each manifest variable. In other words, it only has an intercept factor with freely estimated mean but zero variance, no slope factor, no covariances, and freely estimated error variances. The latter model becomes more stringent by constraining the residual variances in the former model to be equal across time. Based on the work of Widaman and Thompson (2003), Leite and Stapleton (2011) proposed an equivalent null model for their evaluation of single-level latent growth curve models – the outcome means are constrained to be equal across time, covariance are constrained to be zero, but variances are freely estimated.

2.2.3 Selected fit indices

A set of structural equation modeling fit indices are discussed in this section, namely, they are Normed fit index (*NFI*), Tucker-Lewis index (*TLI*), comparative fit index (*CFI*), McDonald's centrality index (*Mc*), root-mean-square error of approximation (*RMSEA*), and standardized root mean squared residual (*SRMR*). These fit indices were selected to cover as many categories of the fit indices discussed in the previous section as possible, with consideration of their practicality and citation rate in simulation and application analyses for both single-level and multilevel-level structural equation models.

Normed Fit Index (NFI)

With T_b and T_h referring to the χ^2 test statistic for the baseline model and the hypothesized model respectively, and with df_b and df_h denoting the degrees of freedom for the baseline model and the hypothesized model respectively, the normed fit index is defined as

$$NFI = \frac{T_b - T_h}{T_b} \tag{13}$$

NFI is the simplest relative fit index, indicating how much a model improves the goodness of fit from the independence model by directly comparing the chi-square statistics of the two models (Sun, 2005). Its values range from 0 to 1, with larger values implying better model fit. Although the theoretical boundary of *NFI* is one, it may never reach this upper limit even if the specified model is correct, especially in small samples (Bentler, 1990). This is because the expected value of the chi-square for the target model which is approximately the degrees of freedom when the model is “true” is always greater than zero (i.e., $E\left(\chi_t^2\right) \approx df > 0$). With small sample size, moreover, *NFI* tends to over-reject the true population model (Fan, Thompson, & Wang, 1999).

Tucker-Lewis Index (TLI)

In contrast to *NFI*, the Tucker-Lewis index controls for model complexity by using $\frac{\chi^2}{df}$ rather than the pure chi-square statistic, as is defined by

$$TLI = \frac{T_b / df_b - T_h / df_h}{T_b / df_b - 1} \tag{14}$$

Because the expected value of $\frac{\chi^2}{df}$ approximately equals one for a “true” model, this index compares the hypothesized model with a “true” model in the ability to correct the chi-square deviation from the independence model (Sun, 2005). *TLI* is only approximately normed in that it can fall out the zero-to-one range. Its value tends toward one for a correctly specified model and a higher value indicates a better model fit. With small sample size, this index is underestimated (Anderson & Gerbing, 1984; Fan, Thompson, & Wang, 1999) and has large sampling variability (Anderson & Gerbing, 1984; Bentler, 1990), which becomes more severe when the sample size decreases. Overall, this index rewards more parsimonious models and penalizes more complex models (Schermelleh-Engel, Moosbrugger, & Muller, 2003).

Comparative Fit Index (CFI)

Being a population-based relative fit index, the comparative fit index indicates how much the hypothesized model corrects the noncentrality of the chi-square distribution from the independence model (Sun, 2005). Given as

$$CFI = 1 - \frac{\max\left[\left(T_h - df_h\right), 0\right]}{\max\left[\left(T_h - df_h\right), \left(T_b - df_b\right), 0\right]} \tag{15}$$

CFI employs the maximum function to strictly bound its values by zero and one in finite samples (Ryu, 2008), with higher values indicating a better model fit. When compared with *TLI*, *CFI* has smaller sampling variability (Bentler, 1990; Yu, 2002) but is still influenced by downward bias when the sample size is small (Fan, Thompson, & Wang, 1999).

McDonald's Centrality Index (Mc)

McDonald's centrality index compares the reproduced covariance matrix with the population covariance matrix instead of the sample covariance matrix (Sun, 2005), as is defined by,

$$M_c = \exp\left[-\frac{1}{2\left(\dfrac{T_h - df_h}{N-1}\right)} \right] \tag{16}$$

with *N* representing sample size[8]. When a model is true or saturated, the chi-square value is approaching its degrees of freedom, making the population discrepancy function close to zero and the *Mc* index approach one. In contrast, when a model is

[8] Some researchers use *N* instead of (*N*-1) to calculate *Mc* (see Sun, 2005).

misspecified, the discrepancy between the chi-square value and its degrees of freedom increases, leading the *Mc* index to approach zero. Hence the *Mc* index typically lies within the range of zero and one (but can exceed one), with larger values indicating better model fit. With small sample size, however, *Mc* tends to depart substantially from its true population values (Hu & Bentler, 1998).

Root Mean Square Error of Approximation (RMSEA)

Based on the population noncentrality parameter, the root mean square error of approximation is defined as[9]

$$RMSEA = \sqrt{\max\left[\frac{T_h - df_h}{df_h\left(N-1\right)}, 0\right]} \tag{17}$$

where the chi-square value for the target model is related to the discrepancy between the population mean and covariance matrices and the model-implied mean and covariance matrices (Wu, 2008). Hence, *RMSEA* can be viewed as a measure of the average discrepancy between the population and model-implied mean and covariance matrices per degree of freedom with the model complexity taken into account (Browne & Cudeck, 1993; Hsu, 2009; Steiger, 2007; Wu, 2008). By applying the maximum function, *RMSEA* is bounded by zero (i.e., meaning that the hypothesized model fits the data perfectly), and a smaller value indicates a better model fit. An advantage of this index is that *RMSEA* has a known sampling distribution, which consequently allows for the calculation of the confidence intervals for this index (Sun, 2005; Yu, 2002). In addition, it favors parsimonious models, but tends to over-

[9] Some researchers use *N* instead of (*N*-1) to calculate *RMSEA* (see Sun, 2005).

reject true models when the sample size is small (i.e., N ≤ 250)[10] (Hu & Bentler, 1998; Iacobucci, 2009).

Standardized Root Mean Square Residual (SRMR)

SRMR can be obtained by using the following formula,

$$SRMR = \sqrt{\frac{2\sum_{i=1}^{p}\sum_{j=1}^{i}\left(\frac{s_{ij}-\hat{\sigma}_{ij}}{s_{ii}s_{jj}}\right)^{2}}{p\left(p+1\right)}} \tag{18}$$

where s_{ij} is a sample covariance between variables *i* and *j*, $\hat{\sigma}_{ij}$ is the corresponding model-implied covariance between variables *i* and *j*, s_{ii} and s_{jj} are the sample standard deviations for the variables *i* and *j* respectively, and *p* is the total number of variables in the model for analysis (Bentler, 1995). The discrepancy between the sample covariance and the corresponding model-implied covariance ($s_{ij} - \hat{\sigma}_{ij}$) indicates the degree of misfit. For multilevel confirmatory structural equation models, this index is calculated separately for both the between-cluster and within-cluster levels. Being a badness-of-fit index in the standardized metric, *SRMR* ranges from zero to one, with zero implying perfect model-data match.

Properties associated with the fit indices are discussed in the context of single-level confirmatory factor analytical models, which may or may not be extended to multilevel models. One index that stands out when compared to other fit indices included in the current study is *SRMR*, which is able to detect misspecification at

---

[10] This statement is in contrast to the finding by Browne and Cudeck (1993) which asserts that *RMSEA* is relatively insensitive to sample size.

different levels, particularly those at the between-level. Moreover, its ability to detect misspecification at the within-level is largely determined by sample size, and its ability to detect misspecification at the between-level declines with the decrease of the intraclass correlation coefficient. More features of the fit indices are summarized in Table 2.

Table 2: Properties of the Fit Indices

		Sensitive To									
		Factor Covariance(s)			**Sample Size**			**Severity Level**			
Fit Index	**Normed**	**CFA**	**MCFA**	**LGC**	**CFA**	**MCFA**	**LGC**	**CFA**	**MCFA**	**LGC**	**ICC**
NFI	Yes	Yes			Yes			Yes			
TLI	No	Yes		Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
CFI	Yes	Yes		Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Mc	No	Yes			Yes			Yes		Yes	
RMSEA	No	Yes		Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
SRMR	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

2.2.4 Cutoff values

One goal in the study of practical fit indices is to develop a set of fixed cutoff values which can work as a criterion for model selection. That is, if the critical value is exceeded in the correct direction, the model is rejected; otherwise, it is accepted. However, this "rule of thumb" is difficult to develop because fit indices are not comparable as they are constructed on different discrepancy functions, they have different sensitivity to sample size, model complexity, and data distribution, and, to some extent, they are prohibited by the inherent inability of a structural model to exactly account for the phenomena it seeks to describe (Sivo, Fan, Witta, & Willse, 2006; Yu, 2002).

Despite of these complexities, researchers have recommended different cutoff values based on particular statistical models (e.g., Browne & Cudeck, 1993; Chen, Curran, Bollen, Kirby, & Paxton, 2008; Hsu, 2009; Hu & Bentler, 1999; Yu, 2002). Among all the cutoff values that have been suggested, the work conducted by Hu and Bentler (1999) is often cited in real practice and simulation studies (see Table 3).

Table 3: Cutoff Values for the Fit Indices Based On Multivariate Normal Outcome

	CFA		**MCFA**	**LGC**
Fit Index	(Hu & Bentler, 1999)	(Yu, 2002)	(Hsu, 2009)	(Yu, 2002)
NFI	≥ 0.90			
TLI	≥ 0.95	≥ 0.95		≥ 0.95
CFI	≥ 0.95	≥0.96	≥ 0.97	≥ 0.95
RMSEA	≤ 0.06	≤ 0.05	≤ 0.054	≤ 0.06
SRMR	≤ 0.08	≤ 0.07	≤ 0.044 (b) ≤ 0.052 (w)	≤ 0.07

In addition to those values, Hu and Bentler also proposed a two-index presentation strategy to control the sum of Type I and Type II error rates. To be more specific, a combination of *SRMR* > 0.09 with a cutoff value of 0.96 for *TLI* and *CFI*, or a combination of *RMSEA* > 0.06 and *SRMR* > 0.09 result in the least sum of Type I and Type II error rates; a combination of *RMSEA* > 0.05 and *SRMR* > 0.06 result in acceptable Type II error rates for simple and complex misspecified models under both robustness and non-robustness conditions. When sample size is small (i.e., $N < 250$), the combination of *CFI* and *SRMR* are more preferable because combinational rules based on *RMSEA* (or *TLI*) and *SRMR* tend to reject more simple and complex true-population models under the non-robustness condition. Furthermore, using combinational rules with *Mc* < 0.90 and *SRMR* > 0.09 yield minimum sum of Type I and Type II error rates. Combinational rules with *Mc* < 0.90 and *SRMR* < 0.06 result in acceptable proportions of simple and complex misspecified models under both robustness and non-robustness conditions. In contrast, when sample size is small, any chosen combinational rules with *Mc* tend to yield relatively large Type I error rates under both robustness and non-robustness conditions. In general, when sample size is small, most of the combinational rules have a slight tendency to over-reject true population models under non-robustness condition and combinational rules with the ML-based *TLI*, *Mc*, and *RMSEA* are less preferable.

Although specific suggestions on the cutoff values of the fit indices have been proposed, the extent to which they are generalizable beyond the correct and misspecified model conditions used in specific studies is questionable (Sivo et al., 2006). Actually, given the population covariance matrix and the model structure, the mean values as well as the distribution of fit indices change with the sample size, the distribution of the data as well as the chosen statistic (Hsu, 2009; Saris, Satorra, & van der Veld, 2009). Those fit indices offer no protection from parameter values unrelated with the misspecification of the model (Saris & Satorra, 1988), and whether a misspecification is detected or not depends heavily on characteristics unrelated to the misspecification itself such as sample size, values of the parameters, number of indicators, etc (Saris, Satorra, & van der Veld, 2009). Thus, cutoff values for fit indices, confidence intervals for model fit/misfit, and power analysis based on fit indices are open to question (Yuan, 2005).

2.2.5 Related research on the performance of the practical fit indices

Hitherto, the most influential studies in evaluating fit indices in selecting structural equation models are those conducted by Hu and Bentler (1998, 1999), which evaluated the sensitivity of maximum likelihood (ML)-, generalized least squares (GLS)-, and asymptotic distribution-free (ADF)-based fit indices (i.e., *NFI*, *BL*86, *TLI*, *BL*89, *RNI*, *CFI*, *GFI*, *AGFI*, Gamma hat, *CAK*, *CK*, *Mc*, *CN*, *SRMR*, and *RMSEA*) to model misspecification in the covariance structure in confirmatory factor analytic models, under conditions that varied sample size (i.e., 150, 250, 500, 1000, 2500, and 5000) and distribution (i.e., the “first was a baseline distributional condition involving normality, the next three involved nonnormal variables that were independently distributed when uncorrelated, and the final three distributional conditions involved nonnormal variables that, although uncorrelated, remained dependent” (Hu & Bentler, 1998, pp. 432)). Their results indicated that (1) most of the ML-based fit indices outperformed those obtained from GLS and ADF estimation method, and should be preferred indicators for model evaluation and selection; (2) *NFI*, *BL*86, *CAK*, *CK*, *CN*, *GFI*, and *AGFI* performed poorly and were not recommended for evaluating model fit; (3) *SRMR* was the most sensitive index to underparameterized factor covariance(s) while *TLI*, *BL*89, *RNI*, *CFI*, Gamma Hat, *Mc*, and *RMSEA* were the most sensitive indices to underparameterized factor loadings; and (4) *NFI*, *BL*86, *GFI*, *AGFI*, *CAK*, and *CK* behaved similarly in model evaluation along the three dimensions discussed in the study (i.e., sample size, distribution, and model misspecification) whereas *TLI*, *BL*89, *RNI*, *CFI*, *Mc*, and *RMSEA* performed similarly in model evaluation on those aspects. Since *SRMR* performed least similarly to all other fit indices, a two-index presentation strategy that using *SRMR*, supplemented with the *TLI*, *BL*89, *RNI*, *CFI*, Gamma Hat, *Mc*, or *RMSEA* was proposed to detect model misspecification in factor covariance(s), factor loading(s), or both. Cutoff values for the two-index presentation strategy were subsequently proposed, which stated that a cutoff value close to 0.95 for *TLI*, *BL*89, *CFI*, *RNI*, and Gamma Hat, a cutoff value close to 0.90 for *Mc*, a cutoff value close to 0.08 for *SRMR*, and a cutoff value close to 0.06 for *RMSEA* were needed for the ML method. In addition, *TLI*, *Mc*, and *RMSEA* tended to over-reject true population models at small sample size and thus were less preferable if using the proposed cutoff criteria.

Fan and Sivo (2005) reevaluated the validity of the two-index presentation strategy and disclosed two design defects associated with Hu and Bentler's studies: (1) the severity of model misspecification was confounded with types of misspecification, which compromised the internal validity of the conclusion that those fit indices were differentially sensitive to different types of model misspecification, and (2) there was an obvious lack of diversity in terms of the models and model parameters examined (i.e., a large number of covariances in the model-based covariance matrix were forced to be zeros), thus jeopardizing the external validity / generalizability of the conclusion that SRMR was the most sensitive index to misspecified factor covariances in general. Consequently, they partially replicated the Hu and Bentler's study by controlling the following five factors: estimation method (i.e., ML versus GLS), type of misspecification (i.e., underparameterization in factor covariance(s) or that in factor loadings), level of misspecification (i.e., true model with no misspecification, a model with one parameter misspecified, and a model with two parameters misspecified), severity of misfit (i.e., slight and moderate as defined by the Satorra-Saris approach based on a sample size of 100), and sample size (i.e., 150 to 1500 at an interval of 150). The results indicated that (1) there was insufficient evidence to support the multifactor view for the fit indices; (2) *SRMR* was not generally most sensitive to misspecified factor covariances and *TLI*, *BL*89, *RNI*, *CFI*, Gamma hat, *Mc*, or *RMSEA* were not more sensitive to misspecified factor loadings. In other words, the validity of the rationale for the proposed two-index strategy no longer held.

In order to address the second deficiency of Hu and Bentler's study, Fan and Sivo (2007) extended the CFA models to have five different types of covariance structures and examined the sensitivity of twelve fit indices (i.e., *NFI*, Rho1, *TLI*, Delta2, *RNI*, *CFI*, *GFI*, *AGFI*, Gamma hat, *Mc*, *SRMR*, and *RMSEA*) to model misspecification when controlling for the severity of misfit. By varying the severity of misfit at two levels (i.e., moderate and severe as defined by the Satorra-Saris approach with a sample size of 100) and sample size at 10 levels (i.e., 100 to 1000 at the interval of 100), they found that (1) absolute fit indices, such as gamma hat, *RMSEA*, or *Mc*, outperformed the incremental fit indices in detecting model misspecification. In addition, among the three absolute fit indices, gamma hat behaved superiorly when model size was small; (2) the sensitivity of most fit indices, including *RMSEA*, *TLI*, *CFI*, and *SRMR* to model misspecification depended on model type, thus making a general cutoff criterion infeasible.

Afterwards, Hu and Bentler's study (1998) was extended to include more complex models, among which multilevel structural equation models and latent growth curve models are most related to the current study. Limited amount of research has been conducted based on those models, and three of them are briefly reviewed in this section.

Hsu (2009) conducted two Monte Carlo studies to investigate the sensitivity of four fit indices (i.e., *RMSEA*, *CFI*, *SRMR-W*/*SRMR-B*, and *WRMR*) in detecting model misspecification in multilevel structural equation modeling – one with normally distributed outcomes while the other with dichotomously distributed outcomes. The design factors included in the first study were number of groups (i.e., N_g = 150, 200, and 250), group size (i.e., N_{gi} = 15 and 30), intraclass correlation coefficient (i.e., 0.40 and 0.18), and model misspecification (i.e., under-parameterized factor covariance and under-parameterized path coefficient). The simulation results showed that *SRMR* was the only index that could detect misspecifications at different levels. In addition, it was more sensitive to the structural model misspecification than the measurement model misspecification for the within-group model, when controlling for other factors. Moreover, it was less likely to detect between-model misspecifications when the *ICC* value decreased. Hence the author suggested using *SRMR* in combination with *RMSEA* and *CFI* to evaluate the within-group model misspecification.

Leite and Stapleton (2011) compared the performance of the likelihood ratio test and six other practical fit indices (i.e., *CFI*, *TLI*, *IFI*, *RMSEA*, *SRMR*, and *Mc*) in detecting misspecifications of growth shape in latent growth curve modeling. Three factors were manipulated in their study: growth shape (i.e., quadratic, plateau, and piecewise), sample size (i.e., 100, 200, 500, 1000, and 2000), and severity level of misfit as calculated by the Satorra-Saris approach with a sample size of 100 (i.e., 0.4, 0.7, and 0.9). The results suggested that (1) the likelihood ratio test performed very well in identifying misspecifications with growth shape; (2) *CFI*, *TLI*, *IFI*, *RMSEA*, and *Mc* had similar sensitivities to the changes in the severity of misspecification, to which *SRMR* was somewhat less sensitive; (3) both *RMSEA* and *Mc* were insensitive to the variation in the population nonlinear growth shape; (4) compared with other indices, *SRMR* was more sensitive to sample size; and (5) the cutoff criteria suggested by Hu and Bentler (1999) did not work well for the detection of misspecified growth shapes in linear growth models.

Wu and West (2010) took another perspective to investigate the sensitivity of the likelihood ratio test as well as four practical fit indices (i.e., *RMSEA*, *SRMR*, *CFI*, and *TLI*) in error detection for latent growth curve modeling. By maneuvering the misspecification to occur with either the marginal mean structure (i.e., constraining the mean quadratic parameter to be zero), or one of the four covariance structures (i.e., constraining the variance of the quadratic slope, the covariance between the intercept and linear slope, and the autoregressive coefficient among the residuals to be zero, and constraining the residual variances to be equal across time), or both the marginal mean and the covariance structures, and by controlling for sample size (i.e., N = 125, 250, 500, and 1000) and severity level (i.e., the one defined by the true model fixed likelihood ratio test statistic based on a sample size of 250, with the values being 0.6, 0.8, and 1.0 respectively), the authors declared that (1) the five fit indices were differentially sensitive to various types of misspecification in the growth curve model even when the severity of misfit was carefully controlled; (2) no fit index was always more (or less) sensitive to misspecification in the marginal mean structure relative to those in the covariance structure; (3) *RMSEA*, *CFI*, and *TLI* were more sensitive to the examined misspecifications than T_{ML} and *SRMR* (except for the covariance between the intercept and the linear slope factors), potentially making them better fit indices; (4) *RMSEA*, *CFI*, and *TLI* were not sensitive to sample size whereas T_{ML} was highly affected by sample size and *SRMR* was affected by sample size under some conditions; (5) saturating the covariance structure substantially improved the sensitivity of the practical fit indices to misspecification in the marginal mean structure. In contrast, saturating the marginal mean structure did not change the sensitivity of the fit indices to misspecification in the covariance structure except for *SRMR*; (6) only *RMSEA* and *CFI* were affected by the interactional effect between misspecification in the marginal mean structure and the covariance structures.

In view of the cutoff values, Sivo, Fan, Witta, and Willse (2006) partially replicated the study of Hu and Bentler (1999) to determine whether the cutoff criteria varied to the true population model (i.e., whether the cutoff criteria depends on the population model). By inheriting the configurations in Hu and Bentler's study, they compared the performance of thirteen practical fit indices (i.e., *GFI*, *AGFI*, *CFI*, *TLI*, *NFI*, *Mc*, Rho1, Delta2, *PGFI*, *PNFI*, *RMR*, *SRMR*, and *RMSEA*), with the purpose to figure out the one(s) that performed optimally across sample size and data distribution. The research results implied that (1) except for *PGFI*, *PNFI* and perhaps *SRMR*, for all the fit indices examined in the study, the cutoff criteria did not vary depending on which model served as the correct model; (2) for correct models, the optimal cutoff values depended on sample size, with smaller sample size resulting in lower optimal cutoff values; (3) for misspecified models, the cutoff values decreased for *Mc*, *SRMR*, and *RMSEA* as sample size increased; and (4) *Mc*, *SRMR*, and *RMSEA* showed a more obvious mean index value discrepancy between correct and incorrect models, suggesting that these indices may do the best job in detecting model misspecification.

Hsu (2009) extended Hu and Bentler's model specification to a multilevel structural equation model, based on which he evaluated the cutoff values for *RMSEA*, *CFI*, *SRMR*, and *WRMR*, with the last one targeting at models for dichotomous outcomes. By fixing the severity of misspecification to 0.70[11], the author claimed that the cutoff values suggested by Hu and Bentler in general resulted in very low statistical powers (i.e., below 0.55) for *RMSEA*, *SRMR*, and *CFI* when the outcomes

[11] The severity level was calculated by the Satorra-Saris approach.

were continuous and multivariate normal. Hence lower cutoff values for *RMSEA* and *SRMR* and higher cutoff value for *CFI* were required.

Yu (2002) evaluated the adequacy of the cutoff criteria by using a quadratic latent growth curve model. By manipulating model misspecification (i.e., quadratic versus linear latent growth curve models), number of time points (i.e., five versus eight time points), sample size (i.e., 100, 250, 500, and 1000), and estimation method (i.e., the ML and the Satorra-Bentler's method), he examined the cutoff values for *TLI*, *CFI*, *RMSEA*, *SRMR* and *WRMR* and concluded that (1) the cutoff criteria proposed by Hu and Bentler were generally suitable for growth models when sample size was no smaller than 250. With smaller sample size (i.e., $N = 100$) and fewer time points (i.e., 5 time points), however, *TLI*, *CFI* and *RMSEA* under the suggested cutoff values tended to over-reject true models; (2) with more time points, *TLI*, *CFI* and *RMSEA* performed well across all sample sizes with the suggested cutoff criteria. In contrast, *SRMR* and *WRMR* tended to over-reject true models; and (3) different cutoff values seemed to be necessary for *WRMR* in latent growth curve modeling with different time points, which might not be desirable.

In sum, the performance of the practical fit indices has been compared based on a variety of CFA models, multilevel SEM models, and latent growth curve models. By controlling factors such as sample size, severity level of misfit, estimation method, intraclass correlation coefficient, and data distribution, researchers have been trying to figure out the fit index(es) that work optimally in model evaluation. Different cutoff values are then proposed and evaluated, which however, have not achieved a unified conclusion.

2.3 Factors Influencing the Evaluation of Linear-Linear Piecewise Multilevel Latent Growth Curve Models

As pointed out by Hu and Bentler (1998), there are four major problems in using fit indices to evaluate model goodness-of-fit: sensitivity of a fit index to model misspecification, small sample bias, estimation method effect, and effects of violation of normality and independence assumptions. Because fit indices are typically based on chi-square tests and the adequacy of a chi-square test statistic depends on sample size and the particular assumptions of a statistical model, these same factors are expected to influence the evaluation of model fit.

Over the past 30 years, methodologists have investigated several data and analytic conditions that influence fit statistics and indices. These include, but are not limited to: type of model, severity of misspecification, sample size, cluster balance, number of variables, estimation method, and indicator reliability (e.g., Anderson & Gerbing, 1984; Bearden, Subhash, & Teel, 1982; Beauducel & Wittmann, 2005; Breivik, & Olsson, 2001; Chen et al., 2008; Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007; Ding, Velicer, & Harlow, 1995; Enders & Finney, 2003; Fan & Sivo, 2005, 2007; Fan, Thompson, & Wang, 1999; Gerbing & Anderson, 1993; Hox & Maas, 2001; Hu & Bentler, 1998, 1999; Kenny & McCoach, 2003; Marsh, Balla, & Hau, 1996; Marsh, Balla, & McDonald, 1988; Sivo et al., 2006; Sugawara & MacCallum, 1993; Taylor, 2008). In terms of the evaluation of multilevel structural equation modeling when the multivariate normality and independence assumptions are met, factors such as model misspecification, severity of misspecification, estimation method, sample size, and intraclass correlation coefficient are explored in most studies. In view of piecewise linear latent growth curve analysis, types of longitudinal data (i.e., data missingness), number of measurement occasions, knot location, and the coding of the time variable have been preliminarily investigated.

2.3.1 Model misspecification

According to Hu and Bentler (1998), there are two types of model misspecifications: one is over-parameterization whereas the other is under-parameterization. The former refers to the condition that “one or more parameters are estimated whose population values are zeros” and the latter refers to the case that “one or more parameters are fixed to zeros whose population values are non-zeros” (pp. 434). In view of structural equation modeling, over-parameterization means saturating the model, which consequently leads to better model fit and improvement in the performance of the fit indices. Under-parameterization, however, deletes information included in a model and consequently undermines the model accuracy. This supposition is supported by the study conducted by La Du and Tanaka (1989) whose results suggest that over-parameterization does not have a significant effect on NFI and GFI whereas under-parameterization has a very small but significant effect on those two fit indices.

For a linear-linear piecewise multilevel latent growth curve model specified in the previous section, there are four potential sources of model misspecification: the marginal mean structure (i.e., $\boldsymbol{\alpha}$), the between-group covariance structure (i.e., $\boldsymbol{\psi}_B$), the within-group covariance structure (i.e., $\boldsymbol{\psi}_W$), and the within-group residual structure (i.e., $\boldsymbol{\Theta}$). Given the assumption that the between-group components are orthogonal and additive to the within-group components (Asparouhov & Muthén, 2007; Preacher, Zyphur, & Zhang, 2010; Searle, Casella & McCulloch, 1992), there are another two potential sources of misspecifications for the specific model, which are the interaction between the marginal mean structure and the between-level covariance structure (i.e., α and ψ_B) and that between the within-level covariance and within-level residual structures (i.e., ψ_W and Θ).

In view of the relationship between the fit of the marginal mean structure and that of the covariance structure, the calculation of the covariance structure is based on the discrepancy between the observed and estimated marginal means such that the magnitude of the discrepancy in the covariance structure increases as the marginal means become increasingly misspecified. In contrast, the marginal means are a function of fixed effects, which are far less sensitive to the specification of the covariance structure. Actually, for linear models, the marginal means are asymptotically independent of the covariance structure for longitudinal data which are complete and balanced on time if the standard assumptions are met (Verbeke & Lesaffre, 1997; Yuan & Bentler, 2004). For realistic sample size, however, the fit of the marginal mean structure is still affected by the covariance structure because the residuals in means are weighted by the estimated covariance structure (Wu, West, & Taylor, 2009).

A challenging issue involving the evaluation of a linear-linear piecewise multilevel latent growth curve model is to differentiate different sources of misspecification. Currently, two approaches are usually adopted to address the issue. One strategy is to utilize a combination of fit indices to reflect different sources of misfit and provide a fuller picture of the adequacy of model fit; and the other is to saturate either the mean or the covariance structure so that the influence of the saturated structure is minimized. However, the extent to which each of these strategies can detect various forms of misspecification still awaits future research (Wu, West, & Taylor, 2009).

2.3.2 Severity of misfit

Quantifying the severity of model misspecification is a tricky issue, as it is usually confounded by types of model misspecification in previous studies on the evaluation of structural equation models (e.g., Fan, Thompson, & Wang, 1999; Hsu, 2009; Hu & Bentler, 1998; Yu, 2002). This consequently undermines the credibility of those studies because meaningful comparisons can only be made across models when the degree of misspecification (e.g., number, type, and magnitude of misspecified parameters) is constant across different levels of other design factors (Fan & Sivo, 2005, 2007). To address this issue, researchers borrowed the idea of the Satorra-Saris approach for power estimation (Satorra & Saris, 1985; Saris & Satorra, 1993) into model evaluation and operationalized the misspecification severity as the power to detect the overall discrepancy of a target model from the population model (Enders & Finney, 2003; Fan & Sivo, 2005, 2007; Leite & Stapleton, 2011; Taylor, 2008; Wu & West, 2010).

The Satorra-Saris approach, initially proposed to address misspecifications concerning only the covariance structure and later extended by Muthén and Curran (1997) to include both the mean and the covariance structures, is said to be desired for very specific model misspecification and particularly suitable for the intervention setting. It employs non-central chi-square distributions, and in particular, the noncentrality parameter (i.e., $\lambda = (N-1)\hat{F}_{ML}$), to gauge the discrepancy between the observed and the estimated mean and covariance structures. Given multivariate normality, a large sample size, and proper model specification, $T_{ML} = (N-1)\hat{F}_{ML}$ is distributed asymptotically as a chi-square distribution. When the model is incorrect but not highly misspecified, Satorra and Saris showed that T_{ML} was asymptotically distributed as a non-central chi-square variate with a certain noncentrality parameter such that the parameter represented the rightward shift from a central chi-square distribution. Assuming that the true model misspecification in such circumstances is equal to or greater than the misspecification due to sampling error (MacCallum, Browne, Sugawara., 1996), the noncentrality parameter then represents the lack of fit of a given model in the population (Muthén & Curran, 1997).

This noncentrality parameter is usually obtained by a two-step procedure. In the first step, the estimated "population" mean and covariance matrices are calculated based on a correctly specified model with pre-determined parameter values. In the second step, a more restrictive model (i.e., some model parameters being constrained to zero) is fit to the estimated mean and covariance matrices obtained in the previous step. The T_{ML} value obtained in the second step is used as an estimate of the noncentrality parameter, and the power to reject a misspecified model at a desired alpha level (i.e., usually $\alpha = 0.05$) is calculated by comparing a non-central chi-square distribution defined by the noncentrality parameter and the degrees of freedom of the misspecified model to a central chi-square distribution with the same degrees of freedom at a given sample size (Muthén & Curran, 1997; Satorra & Saris, 1985; Saris & Satorra, 1993; Saris & Sronkhorst, 1984).

This Satorra-Saris approach is reported to perform sufficiently accurate for practical purposes at small sample size such as 100 observations (Curran, 1994; Saris & Satorra, 1993; Muthén & Curran, 1997). However, its accuracy are prohibited by error propagation associated with the full information maximum likelihood estimation method, which means that misspecification in one part of the model will affect estimates for other parameters elsewhere in the model because all parameters in the model other than the misspecified one are freely estimated (Kline, 2012). In other words, a misspecification in one parameter might be manifested as biased estimate(s) of other parameter(s) without substantially changing the overall estimated mean and covariance matrices (Gerbing & Anderson, 1993). Because it is difficult to predict the direction or magnitude of error propagation (Kline, 2012), this approach is criticized by some researchers (e.g., Wu & West, 2010) who proposed to use the true model fixed likelihood ratio test statistic (TMFLR) to measure the severity of model misspecification when the true model is known. This TMFLR statistic is calculated by fitting the misspecified model to the population mean and covariance matrices with the parameters other than the misspecified one fixed at their population values. In this way, misspecification in one parameter is not cancelled out by other parameters and bias leakage is prevented consequently.

2.3.3 Level-specific evaluation method

Current standard approaches to evaluating the goodness of model fit applies the procedure used in a single-level model to multilevel models and simultaneously evaluates all levels of a multilevel model. A potential problem associated with this practice is that it may not be sensitive to the goodness or badness of a multilevel model at the group level, as is supported by simulation studies (Ryu & West, 2009; Yuan & Bentler, 2007). Because the maximum likelihood estimation function weights the within-group and the between-group models differentially depending on their respective sample size and sample size is typically much larger at the individual level than at the group level, the overall chi-square value and fit indices based on the chi-square value are expected to be dominated by the within-group model (Hox, 2002; Ryu & West, 2009; Yuan & Bentler, 2007). Hence level-specific estimation methods have been proposed to address the defects associated with the simultaneous estimation approach.

Yuan and Bentler (2007) recommended a segregating approach in which the covariance structure of a multilevel model was separated into multiple single-level[12] covariance structures which were then evaluated as in conventional covariance structure analysis. Consequently, level-specific information, including parameter estimates, their standard errors, and the fit indices were obtained. Via a real data example and a simulation study, Yuan and Bentler asserted that this segregating approach was superior to the simultaneous estimation approach in detecting model misspecification at both levels. However, the recovery of the parameter estimates and their standard errors produced by multilevel covariance structure analysis and Yuan and Bentler's segregating method has not been studied (Ryu & West, 2009). Moreover, this approach is limited in simulation studies in that it requires a huge

[12] Yuan and Bentler's procedure takes two steps: (1) produce estimates of saturated covariance matrices at each level, and then (2) perform single-level covariance structure analysis at each level with estimated covariance matrices as input.

amount of computer memory to calculate the within- and between-group covariance matrices when the total sample size is large (e.g., *N* is greater than 10000).

In contrast to Yuan and Bentler's method, Ryu and West (2009) utilized saturation for their level-specific model evaluation and proposed a partially saturated model instead. In this approach, the misspecified level is specified as hypothesized while the other level is saturated and simultaneous estimation is conducted at each level (see Table 4). Ryu and West reported that the parameter estimates as well as their standard errors produced by this method were very close to those of the hypothesized model when misspecification occurred at the within level; and their difference was not substantial when misspecification occurred at the between level. In addition, all the chi-square based fit indices performed well in detecting the between-level misspecification. However, as claimed by Hox (2010), this practice is vulnerable to parameter estimate bias because misspecification at one level may affect the estimates of the saturated model at the other level (Hox, 2010; Yuan & Bentler, 2007), which in turn will attenuate the power of a fit index to detect the misspecification (i.e., fit indices that are mostly sensitive to the degree of fit will show a spuriously good fit whereas those that also reflect the parsimony of the model may show a spurious lack of fit) (Hox, 2010).

Table 4: Algebraic Definitions of the Fit Indices via the Partially-Saturated Estimation[13]

Between-Level Model	Within-Level Model
$NFI_{PS\_B} = \dfrac{T_{IB,SW} - T_{HB,SW}}{T_{IB,SW}}$	$NFI_{PS\_W} = \dfrac{T_{IW,SB} - T_{HI,SB}}{T_{IW,SB}}$
$TLI_{PS\_B} = \dfrac{T_{IB,SW} / df_{IB,SW} - T_{HB,SW} / df_{HB,SW}}{T_{IB,SW} / df_{IB,SW} - 1}$	$TLI_{PS\_W} = \dfrac{T_{IW,SB} / df_{IW,SB} - T_{HW,SB} / df_{HW,SB}}{T_{IW,SB} / df_{IW,SB} - 1}$
$CFI_{PS\_B} = 1 - \dfrac{\max\left[\left(T_{HB,SW} - df_{HB,SW}\right), 0\right]}{\max\left[\left(T_{HB,SW} - df_{HB,SW}\right), \left(T_{IB,SW} - df_{IB,SW}\right), 0\right]}$	$CFI_{PS\_W} = 1 - \dfrac{\max\left[\left(T_{HW,SB} - df_{HW,SB}\right), 0\right]}{\max\left[\left(T_{HW,SB} - df_{HW,SB}\right), \left(T_{IW,SB} - df_{IW,SB}\right), 0\right]}$
$Mc_{\_PS\_B} = \exp\left[-\dfrac{1}{2\left(\dfrac{T_{HB,SW} - df_{HB,SW}}{J}\right)}\right]$	$Mc_{\_PS\_W} = \exp\left[-\dfrac{1}{2\left(\dfrac{T_{HW,SB} - df_{HW,SB}}{N}\right)}\right]$
$RMSEA_{PS\_B} = \sqrt{\max\left[\dfrac{T_{HB,SW} - df_{HB,SW}}{df_{HB,SW}\left(J\right)}, 0\right]}$	$RMSEA_{PS\_W} = \sqrt{\max\left[\dfrac{T_{HW,SB} - df_{HW,SB}}{df_{HW,SB}\left(N\right)}, 0\right]}$

[13] It is actually a partially-saturated model which utilizes ML estimation method.

2.3.4 Intraclass correlation coefficient

The magnitude of the intraclass correlation coefficient of a dataset may affect the accuracy of parameter estimates (Goldstein, 1995; Maas & Hox, 2005). Defined as the proportion of the total variance that is attributable to between-groups differences in the multilevel literature, intraclass correlation coefficient is formulated based on a one-way analysis of variance with random effect (i.e., an empty model), where the outcome on the lower level is the dependent variable and the grouping variable is the independent variable (Raudenbush & Bryk, 2002; Lüdtke et al., 2008)[14]:

[14] Besides the traditional definition of the intraclass correlation coefficient, different definitions were proposed for multilevel growth curve models. In a three-level random-intercept growth curve model (i.e., repeated measures nested within students who are further nested within schools, see Siddiqui, Hedeker, Flay, & Hu, 1996), the respective intraclass correlation coefficients for between-cluster (e.g., schools) and within-cluster (i.e., students) subjects are

$$ICC_g = \frac{v_{11}}{v_{11} + u_{11} + \sigma^2}$$

$$ICC_{gi} = \frac{u_{11}}{v_{11} + u_{11} + \sigma^2}$$

where v_{11} and u_{11} are the variances for the intercepts at the group and individual level respectively, and σ^2 is the level-1 variance. In a three-level random intercept and linear slope model (see Raudenbush & Bryk, 2002), the variances between clusters on individual's initial status and liner growth are

$$ICC_{g\_\text{int}} = \frac{v_{11}}{v_{11} + u_{11}}$$

$$ICC_{g\_slop} = \frac{v_{22}}{v_{22} + u_{22}}$$

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{19}$$

where τ^2 is the variance between groups and σ^2 is the variance within groups.

The same idea is borrowed into multilevel structural equation modeling where Muthén (1991) proposed the concepts of item intraclass correlation and factor intraclass correlation by using a one-factor model. The item intraclass correlation coefficient is the same as that in multilevel analysis whereas the factor intraclass correlation coefficient removes measurement errors from both individual and group levels.

Following Muthén's notation, an observed measurement *y* indexed *j* in the vector $\mathbf{y}_{gi}$ can be decomposed into its fixed effects and the products of factor loadings and random effects,

$$y_{gij} = v_j + \lambda_{Bj}\eta_{Bg} + \varepsilon_{Bgj} + \lambda_{Wj}\eta_{Wgi} + \varepsilon_{Wgij} \tag{20}$$

---

where v_{22} and u_{22} are the variances for the group and individual level slope residuals. Then a more generalized definition of the intraclass correlation coefficient was proposed by Anumendem, (2011), which asserts that the cluster-level variance on any regression parameter can be defined as

$$\rho = \frac{v_{11}}{v_{11} + u_{11} + \sigma^2} \qquad when\ P = 1$$

$$\frac{v_{(p+1)(p+1)}}{v_{(p+1)(p+1)} + u_{(p+1)(p+1)}} \qquad when\ P > 1$$

where *p* is the number of random effects allowed.

where **v** is the overall mean, the λs are factor loadings, η_{Bg} is the between-level random effect, η_{Wgi} is the within-level random effect, and ε_{Bgj} and ε_{Wgij} are the residuals for the between and the within levels respectively. Given that the within and between levels are additive and orthogonal to each other, the variance of y_{gij} can be calculated as

$$\begin{aligned}\sigma^2_{ygij} &= \lambda^2_{Bj}\sigma^2_{\eta B} + \sigma^2_{\varepsilon Bj} + \lambda^2_{Wj}\sigma^2_{\eta W} + \sigma^2_{\varepsilon Wj} \\ &= BF + BE + WF + WE\end{aligned} \tag{21}$$

where B and W stand for between and within levels respectively and F and E represent factor and error respectively. Thus the correlation between two individuals i and i' in group g for variable y_j is,

$$\begin{aligned}ICC &= Corr\left(y_{gij}, y_{gi'j}\right) \\ &= \frac{Cov\left(y_{gij}, y_{gi'j}\right)}{\sigma^2_{ygij}} \\ &= \frac{BF + BE}{BF + BE + WF + WE}\end{aligned} \tag{22}$$

Note that BE is often constrained to be zero in mixed-effects analysis (Muthén, 1991), the above formula can be reduced into

$$ICC = \frac{BF}{BF + WF + WE} \tag{23}$$

where (WF+WE) is the total within-group variability. This is in accordance with the definition of intraclass correlation coefficient in hierarchical linear modeling. Following the logic that the variability within a level can be decomposed into two parts – the factor part and the residual part, the true intraclass correlation coefficient (i.e., factor intraclass correlation coefficient) for a given construct is defined as

$$ICC_T = \frac{BF}{BF + WF} \tag{24}$$

It is clear that this factor intraclass correlation coefficient reflects the true hierarchical relationship in a dataset by removing measurement errors from the computation. Literature (Muthén, 1997b) also suggests that a factor intraclass correlation coefficient decreases slightly over time whereas an item intraclass correlation coefficient increases from pretest to posttest as individuals get more familiar with the topics tested. A limitation of this factor intraclass correlation coefficient lies in its strong assumption of measurement invariance, which requires that the same model structure applies to both the between and the within levels, and which in turn limits the practicality of the coefficient and the generalizability of the research findings (Muthén, 2008).

2.3.5 Sample size

The factor of sample size is discussed here because it exerts great influence on the accuracy of the fit indices in two ways. First, the value of chi-square based fit index increases systematically when sample size becomes larger (Bollen, 1989; Sun, 2005). Second, an increased Type I or Type II error rates may be introduced for small sample sizes (e.g., $N < 200$ for CFA models) with which the asymptotic distribution is not well approximated (Bollen, 1989; Sun, 2005). Sample-based absolute fit indices are relatively vulnerable to sample size effect because sample size enters into the calculation of the fit indices directly; population-based absolute fit indices, however, are more robust to sample size effect because they use sample size as the divisor to mitigate the inflation effect. In contrast to absolute fit indices, relative fit indices are less affected by the change in sample size because they compare the target model to a baseline model, both of which share the same sample size.

In multilevel studies(e.g., Bushing, 1993; Cheung & Au, 2005; Duncan, Alpert, & Duncan, 1998; Duncan et al., 1997; Heck, 2001; Hox, 1993; Hox & Mass, 2001; Kaplan & Elliott, 1997a, 1997b; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002; Lüdtke et al., 2008; Mass & Hox, 2005; Muthén, 1994, 1997a, 1997b; Preacher, Zyphur, & Zhang, 2010; Van der Leeden & Bushing, 1994), the asymptotic nature of the maximum likelihood estimation method used in most studies/software requires sufficiently large sample size, especially at the group-level, because the group-level sample size is always much smaller than the individual-level sample size and the group-level sample size is generally more important than the total sample size[15]. Practitioners often choose individual-level sample size from a couple to more than two dozen and the number of groups from more than a dozen to about five hundred. In the specific field of multilevel latent growth curve modeling, Duncan and his fellows (1997) sampled 203 siblings from 435 families and surveyed their usage of alcohol, marijuana, and cigarette over a four-year period to explore the development of substance use among family members; and Muthén (1997a, 1997b) collected mathematics achievement and attitude data from 2488 and 1869 students who were respectively nested within 50 schools to investigate school effects. In addition, Hox and his colleagues (Hox & Maas, 2001; Hox, Maas, & Brinkhuis, 2010) did two

[15] Maas and Hox (2005) claimed that large individual-level sample sizes might partially compensate for a small number of groups, but Cheung & Au (2005) reached a different conclusion, saying that"increasing the individual-level sample size does not necessarily benefit the parameter estimates and their standard errors at the group level. (pp. 615)"

simulation studies to investigate the robustness of multilevel structural equation modeling toward sample size, intraclass correlation coefficient, and estimation method. By manipulating the group size to be 10-20-50 or 5-10-25, and the number of groups to be 50-100-200, the authors found that 1) an individual-level sample size of 10 sufficed for admissible within-level parameter estimates; 2) a group number of 50 produced accurate fixed effects but deflated standard errors and confidence intervals for the parameter estimates when the data were unbalanced (i.e., the number of individuals in a smaller group was about 1/3 of those in a larger group for each group size). Instead, a group number of 100 was required for sufficient accuracy of the model test and confidence intervals for the parameters; and 3) increasing the group-level sample size to 200 helped with the performance of the maximum likelihood estimation with robust chi-squares and standard errors (MLR) when compared to other estimation methods. However, increasing the individual-level sample size had almost no effect.

In view of piecewise latent growth curve modeling (e.g., Diallo & Morin, 2014; Flora, 2008; Kohli & Harring, 2013; Kohli, Harring, & Hancock, 2013; Kwok, Luo, & West, 2010; Liu, Liu, Li, & Zhao, 2015; Sterba, 2014), largely varying individual-level sample sizes are observed, which range from 30 to 3000. Diallo and Morin’s simulation study suggested that 1) a sample size of 30 was sufficient to detect an obvious distinct second phase slope (i.e., μ_{s2}= 0.55) with power levels over 0.80. However, a sample size of 200 was required to detect a slightly distinct second phase slope (i.e., μ_{s2}= 0.11); 2) if there were more than two measurement occasions before the knot, a sample sizes of 1,500 to 2,000 were needed to detect a small slope mean differences (i.e., μ_{s1} - μ_{s2} = 0.05) at a power level of 0.80. However, this number dropped to 200 when detecting a moderate slope mean differences (i.e., μ_{s1} - μ_{s2} = 0.16), and to 30 to 50 when detecting a large slope mean differences (μ_{s1} - μ_{s2} = -0.39); and 3) when there was only two measurement occasions before the turning point, the required sample sizes for the detection of the mean slope difference rose to 2000~3000, 200~300, and 50 respectively.

In addition to individual-level and group-level sample sizes, another aspect of sample size that is discussed in some multilevel structural equation model studies is cluster balance, which means equal number of individuals across groups. Some simulation studies (e.g., Hox & Mass, 2001; Hox, Mass, & Brinkhuis, 2010; Maas & Hox, 2005) showed that there was no discernible effect of unbalancedness on multilevel parameter estimates or on their standard errors if the group size ratio between the large group and the small group was constrained to be 1/3. However, other studies (e.g., Browne, 2006) suggested that cluster unbalancedness influences the power curves of parameter estimates for multilevel models, with the power curve for balanced data hanging above the power curve for slightly or moderately unbalanced data. Moreover, with severe data unbalance, the power curve behaves abnormally and unpredictably. This is because extremely unbalanced designs are really estimating the effect of large groups instead of the global average, thus making the between-group variance to be zero (Browne, 2006).

A simulation study conducted by Muthén and Curran (1997) also confirmed that “the power curves are not completely symmetric around the balanced case where the proportion is 0.5” (i.e., the power curves for balanced and unbalanced data are different) (pp. 388). By conducting a two-group analysis (i.e., treatment group vs. control group), they asserted that it was more favorable to choose observations from the group that has larger variance.

2.3.6 Number of measurements

In view of methodological and substantive research on piecewise growth curve modeling, most studies contain at least six number of measurement occasions (e.g., Bollen & Curran, 2006; Cudeck, 1996; Cudeck & Klebe, 2002; Diallo & Morin, 2014; Duncan, Duncan, & Strycker, 2006; Harring, Cudeck, & du Toit, 2006; Leite & Stapleton, 2011). In structural equation modeling, the choice of a specific functional form is closely related to the number of measurement occasions that is available. Traditional structural equation modeling theory suggests that at least three time points are necessary to identify a basic linear latent growth curve model (Bollen & Curran, 2006). Fan and Fan's study (2005) further demonstrated that an even larger number of time points were needed to avoid the high non-convergence rate (i.e., 40%) associated with the detection of a linear latent growth trajectory. Moreover, an additional slope factor or more complicated residual structure (e.g., heteroscedastic residuals) requires additional time points. As is shown by Bollen and Curran (2006), at least five time points are necessary to identify a two-piece linear growth curve model when the knot is at the third time point. This is because one of the time points (i.e., the knot) is shared by two time segment, thus both pieces satisfy the requirement of having at least three observations.

It has been argued that an increased number of measurement points results in greater precision in the estimation of latent growth curve models (Cheong, 2011;

Singer &Willett, 2003) and lower rates of non-convergence (Diallo & Morin, 2014; Fan & Fan, 2005). Moreover, an increased number of measurement points are documented to slightly decrease the Type I error rates in detecting the mean slope for the second phase and significantly increase the power to detect mean slope differences between two consecutive development phases (Diallo & Morin, 2014). However, the number of measurements does not affect the power to detect linear growth if the model is linear (Fan & Fan, 2005) or the power to detect the second phase slope if the model is linear-linear piecewise (Diallo & Morin, 2014).

2.3.7 Location of knot

A major problem when using a piecewise latent growth curve model is to specify a priori the precise location of the knot(s) where the change(s) in growth rates occur (Crawford, Pentz, Chou, Li, & Dwyer, 2003; Muthén & Muthén, 2012). In some cases, the location of the knot(s) can be theoretically determined (e.g., transition between different phases of schooling, or immediately before or after a treatment/intervention); in other cases, however, they are more exploratory and may be suggested by the examination of lattice and spaghetti plots (Weiss, 2005).

Kohli and Harring (2013) did a simulation study to investigate the influence of knot position on parameter estimates based on a second-order linear-linear piecewise latent growth curve model. By manipulating factors including sample size (i.e., N = 100, 250, and 500), location of knot (i.e., knot was placed at the 2nd, the 4th, and the 6th time point for a time span of 9 repeated measures), and indicator reliability (i.e., 0.45 and 0.85 for poor and good reliability respectively), the authors concluded that the location of knot was systematically related to parameter bias with respect to the estimation of the slope of the second time segment, the variance of the intercept of the first time segment, and the variance of random disturbances in the first-order latent factors. To be more specific, when shifting the knot to later measurement occasions, the slope of the second phase was negatively biased and the variability of the first-phase intercept was positively biased. Moreover, the relative bias for the variance of random disturbances in the first-order latent factors ranged from 30% to 36%.

Diallo and Morin (2014) employed a linear-linear piecewise latent growth curve model to investigate its power to detect nonlinear growth trajectories. By putting the knot at the 2^{nd}, the 3^{rd}, and the 4^{th} time points over a 6-measurement time span, the authors found that 1) the location of the knot did not statistically significantly influence the power and Type I error rates of the model to detect the mean slope of the second phase; 2) the knot occurring at the 3^{rd} or the 4^{th} occasion did not significantly differ from one another in their power to detect the mean difference in growth rates. The knot occurring at the 2^{nd} occasion, however, led to a slight but statistically significant decrease in that power; and 3) the convergence rate of the model was impacted by the number of measurements before the knot but not after the knot. Moreover, the knot location was effective only when the number of measurements before the knot was not fully optimal (i.e., more than 3 measurement occasions).

2.3.8 Data missingness

The advantage of growth curve modeling over ANOVA analysis is that it is able to analyze longitudinal data that are not collected under panel design (i.e., each individual is observed at the same set of fixed time points and no observations are missing). In other words, growth curve modeling is capable of dealing with data missingness. According to Raudenbush (2001), there are two types of data missingness concerning longitudinal data – one is that individual observations are still collected at the same set of fixed time points but are missing at random, which may result from drop out, attrition, or other reasons; and the other is that individual observations are collected at different time points, leading to abundant patterns of missingness. In the first case, the time variable is treated as ordinal; however, it is regarded as continuous in the second case. Because multilevel latent growth curve model treats subjects with different time collection schemes as separate groups, the model specification may become perplexing if the time variable is continuous[16]. Moreover, maximum likelihood estimation may become computationally heavy, if not impossible at all.

In addition to the availability of the model that can be applied to the data, data missingness may lead to biased parameter estimates and their confidence intervals as well as inflated Type I error rate (Davey, Savla, & Luo, 2005). To be more specific, when data are missing completely at random (i.e., the values of unobserved variables do not depend on the values of observed variables or the missing data) (Little & Rubin, 1989, 2002), the analysis of complete cases results in unbiased parameter estimates, but their confidence intervals are unnecessarily large. If data are only missing at random (i.e., the values of unobserved data depend completely on the values of observed variables but not on the missing data), the analysis of complete cases will result in biased parameter estimates as well as confidence intervals. Furthermore, missing data leads to reduced statistical power to

[16] As is stated above, missingness may or may not depend on data collection design.

reject a misspecified model (Davey, Savla, & Luo, 2005). In view of structural equation modeling fit indices, both chi-square based and residual-based fit indices are affected by data missingness. When each individual is observed at different time points, it is impossible to generate a saturated model for chi-square based fit indices due to the lack of homogeneous mean and covariance matrices across individuals (Browne & Arminger, 1995; Raudenbush, 2001; Wu, West, & Taylor, 2009). It is also infeasible to calculate the residual-based fit indices because they require “a common sample and model implied mean and covariance structure” (Wu, West, & Taylor, 2009). When individuals are observed at a set of fixed time points but are missing at random, those missing data greatly reduces the ability of some fit indices (i.e., *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA*) to detect misspecified covariance structures (Davey, Savla, & Luo, 2005). Even when models are correctly specified, *NFI* is found to perform poorly as missing observations increase, with the bias heaping with data missing at random when compared to the case in which data are missing completely at random (Davey, Savla, & Luo, 2005).

2.3.9 Time coding

The effects of different time scales (i.e., changing the basis function by adding or subtracting a constant, or changing the time point of initial level) have been investigated in both multilevel growth curve modeling (i.e., hierarchical linear modeling approach) and latent growth curve modeling. Anumendem (2011) proved both analytically and empirically that different time coding schemes affect the estimates and interpretations of both fixed and random effects of a multilevel growth curve model. In terms of latent growth curve modeling, different growth parameters as well as their variance and covariance matrices will be obtained if different time scales are adopted for the model (Garst, 2000; Mehta & West, 2000; Rogosa & Willett, 1985; Rovine & Molenaar, 1998). Furthermore, for a linear latent growth curve model, the effects of covariates on the growth parameters also depend on the time scales involved. Stoel, van den Wittenboer, and Hox (2004) demonstrated that standard errors and test statistics of some of the parameters of a latent growth curve model change as a consequence of a different scaling of the basis function of the growth rate (i.e., linear transformation of the basis function).

Although it is an important factor, time coding is mostly explored in application studies because it often involves the interpretation of the growth parameters and the exogeneous covariates (e.g., Anumendem, 2011; Stoel & van den Wittenboer, 2003). Given the fact that the maximum likelihood estimation is a scale-free method (Long, 1984), transformations of the time scale (i.e., changing the factor loadings for a latent growth curve model) will be absorbed by the corresponding changes in the factor variance/covariances (Anumendem, 2011), which confounds the sources of misspecification. Hence it is excluded from the current study.

2.4 Research Questions

Given the emerging applications of linear-linear piecewise multilevel latent growth curve models in real practice and the lack of empirical support for the choice of an appropriate model based on the practical fit indices, the current study aims to evaluate the sensitivity of the commonly used practical fit indices in detecting misspecifications in linear-linear piecewise multilevel latent growth curve models. To be more specific, the study tries to answer the following questions:

1. How do those fit indices react to different types of misspecification?
2. What manipulated factors (including types of misspecification, severity level of misfit, sample size, intraclass correlation coefficient, and cluster balance) affect the performance of those fit indices? And how do they influence the performance of the fit indices?
3. Is each fit index equally sensitive to different types of misspecifications?
4. Are those fit indices comparable in detecting one specific type of misspecification?
5. How do the recommended cutoff values work in detecting model misspecification with linear-linear piecewise multilevel latent growth curve models (i.e., the respective Type I and Type II errors associated with the usage of the cutoff values)?
6. Is it possible to suggest a new cutoff value for each fit index to reject a misspecified linear-linear piecewise multilevel latent growth curve model on a certain alpha level?
7. Does partially-saturated estimation method improve the sensitivity of the fit indices to misspecified between-level covariance structure on small sample size?

Chapter 3: Method

3.1 Simulation Design 1

A Monte Carlo study is conducted via *Mplus* 7 to investigate the sensitivity of six commonly used practical fit indices (i.e., *NFI*, *TLI*, *CFI*, *Mc*, *RMSEA*, and *SRMR*) in detecting different types of model misspecification concerning linear-linear piecewise multilevel latent growth curve models. Among all the functions that have been employed to analyze growth trend in real data analysis, piecewise functions have gained much popularity nowadays because it can accommodate to data whose change is discontinuous. In addition, it can provide more nuanced information about both the nature of growth and the predictors of growth during certain segments within the trajectory. Among piecewise growth curve models, linear piecewise growth curve models have drawn much attention to researchers and practitioners due to its prominent flexibility in fitting curves, its comparatively simple structure and its adaptivity to real data (Hindman, Cromley, Skibbe, & Miller, 2011). The current study adopts a linear-linear piecewise multilevel latent growth curve model based on the analyses of the LSAY data, with the consideration of integrating real data in simulation designs so as to address the defects of previous studies on the sensitivity of fit indices (i.e., they usually employ impractically small parameter values). The general population model is presented in a path diagram shown below (see Figure 2).

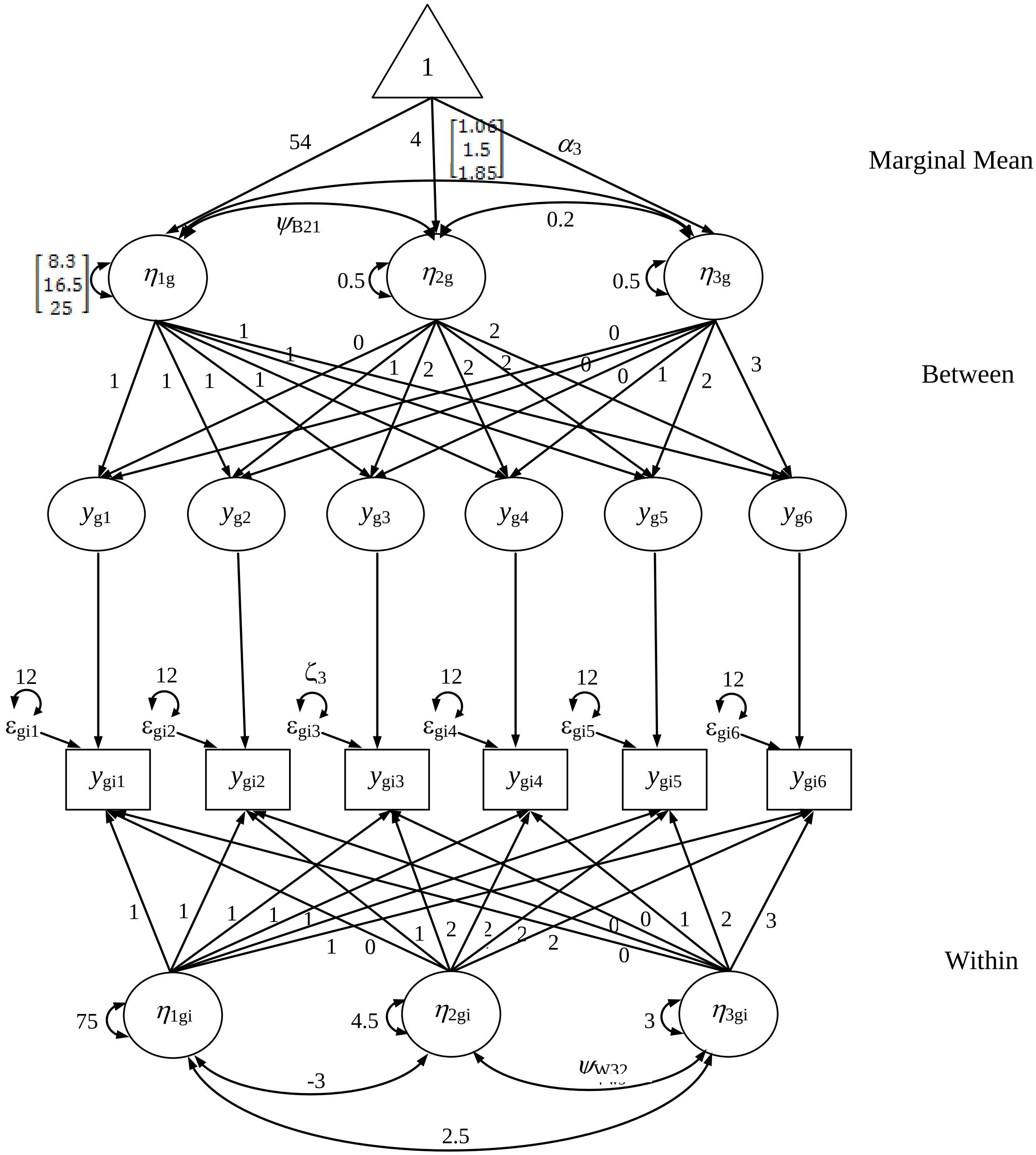


Figure 2: The general population model for linear-linear piecewise multilevel latent growth modeling.

In this model, the number of repeated measurements and the location of the knot are fixed because they are based on the results of the LSAY data. Indicator reliability is also excluded from the design because LSAY test scores are based on the items developed by the National Assessment of Educational Progress (NAEP), which is believed to have high indicator reliability. Minor adjustments are made to the residual covariance structure obtained from the LSAY results in that five out of six residual variances are constrained to be equal. This is manipulated such that the misspecification of the residual structure could be simplified. By fixing the within-level covariance matrix, different levels of the between-level covariance matrix are determined by the intraclass correlation coefficient. Fixed population parameters for that linear-linear piecewise multilevel latent growth curve model are summarized in Table 5.

Table 5: Data Generation Matrix for the Population Models for Multilevel Latent Growth Modeling

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} 54.00 \\ 4.00 \\ \alpha_3 \end{pmatrix}$$

$$\mathbf{\Psi}_{B\_ICC(0.10)} = \begin{pmatrix} 8.30 & & \\ \psi_{B21} & 0.50 & \\ 1.06 & 0.20 & 0.50 \end{pmatrix}, \mathbf{\Psi}_{B\_ICC(0.18)} = \begin{pmatrix} 16.50 & & \\ \psi_{B21} & 0.50 & \\ 1.50 & 0.20 & 0.50 \end{pmatrix}, \mathbf{\Psi}_{B\_ICC(0.25)} = \begin{pmatrix} 25.00 & & \\ \psi_{B21} & 0.50 & \\ 1.85 & 0.20 & 0.50 \end{pmatrix}$$

$$\mathbf{\Psi}_W = \begin{pmatrix} 75.00 & & \\ -3.00 & 4.50 & \\ 2.50 & \psi_{W32} & 3.00 \end{pmatrix}$$

$$\mathbf{\Theta} = \begin{pmatrix} 12.00 & & & & & \\ & 12.00 & & & & \\ & & \zeta_3 & & & \\ & & & 12.00 & & \\ & & & & 12.00 & \\ & & & & & 12.00 \end{pmatrix}$$

The baseline model for incremental fit indices (i.e., *NFI*, *TLI* and *CFI*) is a null model with equal means across time, zero covariances, and freely estimated equal residual variances on both levels (see Figure 3).

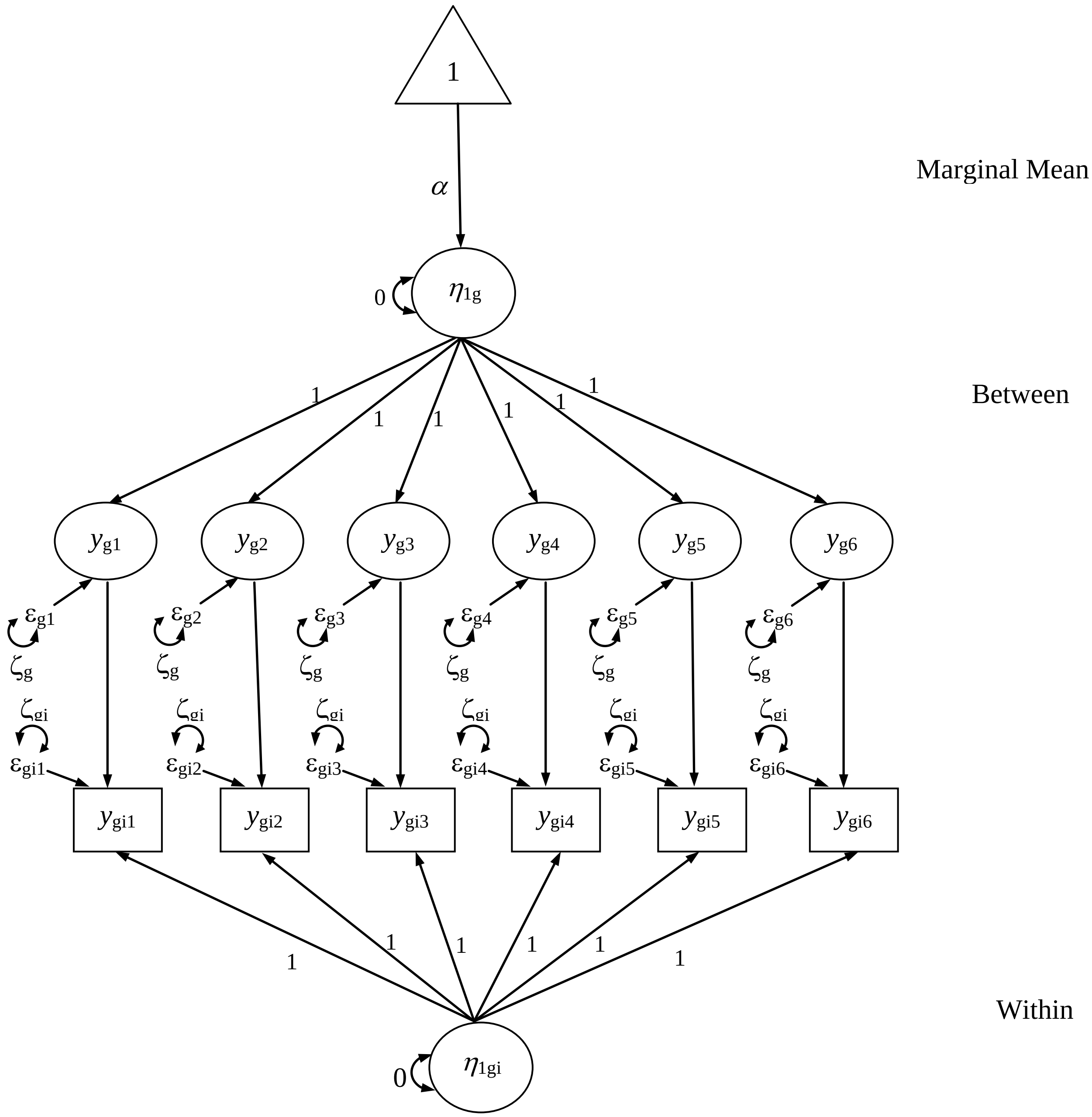


Figure 3: The general baseline model for incremental fit indices.

Given that the number of repeated measures and the number and location of the knot are fixed for the population model, only five factors are manipulated for the current study, which are type of misspecification, severity level of misspecification, sample size (including the individual- and group-level sample sizes), intraclass correlation coefficient, and cluster balance.

Source of misspecification. Tested models are either misspecified in the marginal mean structure, the between-level covariance structure, the within-level covariance structure, the within-level residual structure, or both the marginal mean and the between-level covariance structures, or both the within-level covariance and residual structures. The interactional effect between the marginal mean structure and the between-level covariance structure is included here because the mean structure fell on the between-level for a multilevel latent growth curve model. In addition, no misspecifications across levels are defined because the between-level and within-level are supposed to be independent from each other.

By misspecifying the mean structure, the slope for the second phase (i.e., α_3) of the linear-linear piecewise multilevel latent growth curve model is constrained to the value of the first slope (i.e., fitting a linear function to the data); by misspecifying the covariance structures, the target parameter values (i.e., ψ_{b21} for the between-level covariance structure and ψ_{w32} for the within-level covariance structure) are constrained to be zero; and for misspecifying the residual structure, the structure type is constrained from banded main diagonal to simple (i.e., $\zeta_3 = \zeta_{1, 2, 4, 5, 6}$). Only under-parameterized misspecification is considered in this study because over-parameterized misspecified models have zero population noncentrality and do not have significantly different estimates for model fit indices (La Du & Tanaka, 1989; Hu & Benter, 1999). Moreover, no variances are manipulated in the current study because constraining a variance to zero often influences its covariance with other growth parameters, whose impact could not be carefully calibrated.

Severity of misspecification. Three levels of severity of misfit are included in the study: 0.60, 0.80, and 0.99, which correspond to low, moderate, and high severity levels respectively. The low and moderate values are based on previous research; the high value is obtained by selecting iteratively from the range of [0.90, 1.00] so as to simultaneously obtain as high power and convergence rates as possible. The severity level, as defined by the true model fixed likelihood ratio test statistic, is calculated on a sample of 20 individuals nested with 200 groups. 200 groups are chosen because this is the sample size suggested by Diallo and Morin (2014) to achieve a satisfactory power rate (i.e., higher than 0.80) for the detection of a small mean slope for the second phase in a two-piece linear latent growth curve model.

Sample size. 50, 100, and 200 groups with 10, 20, and 30 individuals nested within each group are employed to detect the effect of sample size on model estimation. The lower bounds for group number (i.e., 50) and group size (i.e., 10) are based on the study conducted by Hox and his fellows (2001, 2010). The larger group numbers (i.e., 100 and 200) are adopted to provide more unique information about the data, and the larger group size takes into consideration the actual average number of students in primary and high schools in the United States. For example, 20 is the average number of students in public primary schools and 30 is the largest average number of students per class in public high schools (National Center for Education

Statistics, 2011-2012). These numbers are considered because previous research on the application of multilevel latent growth curve modeling often uses irregularly large values.

Intraclass correlation coefficient. Few studies report latent variable intraclass correlation coefficient. Instead, the traditional intraclass correlation coefficient is reported for multilevel models concerning educational and organizational studies (e.g., Bliese, 2000; Duncan et al., 1997; Gulliford, Ukoumunne, & Chinn, 1999; Hox & Maas, 2001; James, 1982; Lüdtke et al., 2008; Maas & Hox, 2005; Muthén, 1997a, 1997b; Snijders & Bosker, 1999). In general, there is no 'gold' standard concerning the cutoff value for the intraclass correlation coefficient. Julian (2001) suggests that "when the magnitude of the intraclass correlations are less than 0.05 and the group size is small, the consequences of ignoring the data dependence within multilevel data structures seem to be negligible. … [T]he chi-square statistic is only minimally inflated, and the model parameters and their standard errors are essentially unbiased" (pp. 347). Muthén and Satorra (1995) suggest that even for a rather small intraclass correlation of 0.10, the distortions (i.e., inflated chi-square values and Type I error rate, and deflated standard errors of estimates) may be large if the group size exceeds 15. Actually, Muthén (1997b) claims that an intraclass correlation coefficient of 0.10 paired with group size much less than 50 would jeopardize the maximum likelihood estimator. In any event, the cutoff value of 0.10 may be relatively conservative when deciding the usage of multilevel modeling (Bickel, 2007; Kline, 2011). Thus 0.10 is chosen as the lower boundary for the intraclass correlation coefficients in this study.

Three levels of intraclass correlation coefficient are included in the study: 0.10, 0.18, and 0.25. Among the three values, 0.18 is the value that is obtained from the LSAY scores and 0.25 is the value that is often seen in literature (e.g., Bliese, 2000; Duncan et al., 1997; Gulliford, Ukoumunne, & Chinn, 1999; Hox & Maas, 2001; James, 1982; Lüdtke et al., 2008; Maas & Hox, 2005; Muthén, 1997a, 1997b; Snijders & Bosker, 1999). Although 0.25 is a comparatively low intraclass correlation coefficient in the sense that it is a factor intraclass correlation coefficient value, the value of 0.30 which is often adopted in previous simulation studies is rarely seen in educational organizational studies.

The manipulation of the intraclass correlation coefficient in multilevel structural equation modeling is often realized by changing factor loadings on the between-level while fixing the variance-covariance matrices on both the between and the within levels (e.g., Hsu, 2009). However, as factor loadings indicate the time variable in multilevel latent growth curve models, modifying factor loadings means altering the time scale, which consequently changes the interpretation of the model completely. Hence it is more preferable to manipulate the between-level factor variance to achieve different intraclass correlation coefficient values. Considering the specific population model used in the current study, factor intraclass correlation coefficient instead of item intraclass correlation coefficient is used because the residuals are not constant over time. This configuration of the intraclass correlation coefficient, however, may limit the generalizability of the study results.

Cluster balance. Two levels of cluster balance, balancedness and unbalancedness are designated for the factor. The former means that all the groups have equal number of individuals in each group; and the latter suggests that the number of individuals nested within each group is different. This factor is included in the design because it is closely related to the power curve that is used to obtain the severity level of misfit. With the unbalanced cluster design, the group size ratio between the large group and the small group is set to be 1/3 (i.e., the large group size was three times as large as the small group size), as is specified by Hox and his colleagues (2001, 2010). This small ratio is adopted because extreme cluster unbalancedness results in aberrant growth curves (Browne, 2006), which makes the computation of the severity level of misfit impossible.

The manipulation of the above five factors results in 90 population models, whose values are listed in Tables 6-11. Use "B", "N_g", "N_{gi}", "*ICC*", "S", "T", and "M" to denote cluster balance, group number, group size, intraclass correlation coefficient, severity level, true model, and misspecified models respectively, there are

$B(2)\times N_g(3)\times N_{gi}(3)\times ICC(3)\times T$[17]$(1)$ +

$B(2)\times N_g(3)\times N_{gi}(3)\times ICC(3)\times S(3)\times M(4)$ +

$B(2)\times N_g(3)\times N_{gi}(3)\times ICC(3)\times S(\alpha_3/\psi_{W32})(3)\times S(\psi_{B21}/\zeta_3)(3)\times M(2) = 1674$ (conditions)

included in the study. Data are generated and analyzed for all the 1674 conditions with 1000 replications for each condition.

[17] Population models that are used to detect the misspecification in α_3 work as the true model in this study. Although the severity level is used in data generation, it is not included in data analysis.

Table 6: Values of the Key Parameters in the Population Model and Models Misspecified in the Marginal Mean Structure

				Population Values	
PM	**ICC**	**Severity**	**Misspecification**	**Balanced Data**	**Unbalanced Data**
1	0.10	0.60	$\alpha_3 = 4$	$\alpha_3 = 1.917782835$	$\alpha_3 = 1.918295021$
2	0.10	0.80	$\alpha_3 = 4$	$\alpha_3 = 1.920732835$	$\alpha_3 = 1.921245021$
3	0.10	0.99	$\alpha_3 = 4$	$\alpha_3 = 1.931255835$	$\alpha_3 = 1.931768021$
4	0.18	0.60	$\alpha_3 = 4$	$\alpha_3 = 1.857030546$	$\alpha_3 = 1.857531550$
5	0.18	0.80	$\alpha_3 = 4$	$\alpha_3 = 1.859980546$	$\alpha_3 = 1.860481550$
6	0.18	0.99	$\alpha_3 = 4$	$\alpha_3 = 1.870503546$	$\alpha_3 = 1.871004550$
7	0.25	0.60	$\alpha_3 = 4$	$\alpha_3 = 1.764294412$	$\alpha_3 = 1.764556908$
8	0.25	0.80	$\alpha_3 = 4$	$\alpha_3 = 1.767244412$	$\alpha_3 = 1.767506908$
9	0.25	0.99	$\alpha_3 = 4$	$\alpha_3 = 1.777767412$	$\alpha_3 = 1.778029908$

Table 7: Values of the Key Parameters in the Population Model and Models Misspecified in the Between-Level Covariance Structure

				Population Values	
PM	**ICC**	**Severity**	**Misspecification**	**Balanced Data**	**Unbalanced Data**
10	0.10	0.60	$\psi_{B21} = 0$	$\psi_{B21} = -0.428943994$	$\psi_{B21} = -0.428427570$
11	0.10	0.80	$\psi_{B21} = 0$	$\psi_{B21} = -0.425993994$	$\psi_{B21} = -0.425477570$
12	0.10	0.99	$\psi_{B21} = 0$	$\psi_{B21} = -0.415470994$	$\psi_{B21} = -0.414954570$
13	0.18	0.60	$\psi_{B21} = 0$	$\psi_{B21} = -0.639670778$	$\psi_{B21} = -0.639165506$
14	0.18	0.80	$\psi_{B21} = 0$	$\psi_{B21} = -0.636720778$	$\psi_{B21} = -0.636215506$
15	0.18	0.99	$\psi_{B21} = 0$	$\psi_{B21} = -0.626197778$	$\psi_{B21} = -0.625692506$
16	0.25	0.60	$\psi_{B21} = 0$	$\psi_{B21} = -0.852392799$	$\psi_{B21} = -0.852126045$
17	0.25	0.80	$\psi_{B21} = 0$	$\psi_{B21} = -0.849442799$	$\psi_{B21} = -0.849176045$
18	0.25	0.99	$\psi_{B21} = 0$	$\psi_{B21} = -0.838919799$	$\psi_{B21} = -0.838653045$

Table 8: Values of the Key Parameters in the Population Model and Models Misspecified in the Within-Level Covariance Structure

				Population Values	
PM	**ICC**	**Severity**	**Misspecification**	**Balanced Data**	**Unbalanced Data**
19	0.10	0.60	$\psi_{W32} = 0$	$\psi_{W32} = 0.323122304$	$\psi_{W32} = 0.323718554$
20	0.10	0.80	$\psi_{W32} = 0$	$\psi_{W32} = 0.913118304$	$\psi_{W32} = 0.913714554$
21	0.10	0.99	$\psi_{W32} = 0$	$\psi_{W32} = 3.017832304$	$\psi_{W32} = 3.018428554$
22	0.18	0.60	$\psi_{W32} = 0$	$\psi_{W32} = 0.265931332$	$\psi_{W32} = 0.266526018$
23	0.18	0.80	$\psi_{W32} = 0$	$\psi_{W32} = 0.855927332$	$\psi_{W32} = 0.856522018$
24	0.18	0.99	$\psi_{W32} = 0$	$\psi_{W32} = 2.960641332$	$\psi_{W32} = 2.961236018$
25	0.25	0.60	$\psi_{W32} = 0$	$\psi_{W32} = 0.175530976$	$\psi_{W32} = 0.175891543$
26	0.25	0.80	$\psi_{W32} = 0$	$\psi_{W32} = 0.765526976$	$\psi_{W32} = 0.765887543$
27	0.25	0.99	$\psi_{W32} = 0$	$\psi_{W32} = 2.870240976$	$\psi_{W32} = 2.870601543$

Table 9: Values of the Key Parameters in the Population Model and Models Misspecified in the Within-Level Residual Structure

				Population Values	
PM	**ICC**	**Severity**	**Misspecification**	**Balanced Data**	**Unbalanced Data**
28	0.10	0.60	$\zeta_3 = 12$	$\zeta_3 = 6.651643709$	$\zeta_3 = 6.652016958$
29	0.10	0.80	$\zeta_3 = 12$	$\zeta_3 = 7.241639709$	$\zeta_3 = 7.242012958$
30	0.10	0.99	$\zeta_3 = 12$	$\zeta_3 = 9.346353709$	$\zeta_3 = 9.346726958$
31	0.18	0.60	$\zeta_3 = 12$	$\zeta_3 = 6.589121539$	$\zeta_3 = 6.589475378$
32	0.18	0.80	$\zeta_3 = 12$	$\zeta_3 = 7.179117539$	$\zeta_3 = 7.179471378$
33	0.18	0.99	$\zeta_3 = 12$	$\zeta_3 = 9.283831539$	$\zeta_3 = 9.284185378$
34	0.25	0.60	$\zeta_3 = 12$	$\zeta_3 = 6.495666342$	$\zeta_3 = 6.495779085$
35	0.25	0.80	$\zeta_3 = 12$	$\zeta_3 = 7.085662342$	$\zeta_3 = 7.085775085$
36	0.25	0.99	$\zeta_3 = 12$	$\zeta_3 = 9.190376342$	$\zeta_3 = 9.190489085$

Table 10: Values of the Key Parameters in the Population Model and Models Misspecified in Both the Marginal Mean & the Between-Level Covariance Structures

		Severity		**Population Values**	
PM	**ICC**	(α_3/ψ_{B21})	**Misspecification**	**Balanced Data**	**Unbalanced Data**
37	0.10	0.60/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.919231249$, $\psi_{B21} = -0.428943994$	$\alpha_3 = 1.919709089$, $\psi_{B21} = -0.428427570$
38	0.10	0.80/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.922181249$, $\psi_{B21} = -0.428943994$	$\alpha_3 = 1.922659089$, $\psi_{B21} = -0.428427570$
39	0.10	0.99/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.932704249$, $\psi_{B21} = -0.428943994$	$\alpha_3 = 1.933182089$, $\psi_{B21} = -0.428427570$
40	0.10	0.60/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.919195923$, $\psi_{B21} = -0.425993994$	$\alpha_3 = 1.919674878$, $\psi_{B21} = -0.425477570$
41	0.10	0.80/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.922145923$, $\psi_{B21} = -0.425993994$	$\alpha_3 = 1.922624878$, $\psi_{B21} = -0.425477570$
42	0.10	0.99/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.932668923$, $\psi_{B21} = -0.425993994$	$\alpha_3 = 1.933147878$, $\psi_{B21} = -0.425477570$
43	0.10	0.60/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.919069132$, $\psi_{B21} = -0.415470994$	$\alpha_3 = 1.919551849$, $\psi_{B21} = -0.414954570$
44	0.10	0.80/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.922019132$, $\psi_{B21} = -0.415470994$	$\alpha_3 = 1.922501849$, $\psi_{B21} = -0.414954570$
45	0.10	0.99/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.932542132$, $\psi_{B21} = -0.415470994$	$\alpha_3 = 1.933024849$, $\psi_{B21} = -0.414954570$
46	0.18	0.60/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.860137909$, $\psi_{B21} = -0.639670778$	$\alpha_3 = 1.860598208$, $\psi_{B21} = -0.639165506$
47	0.18	0.80/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.863087909$, $\psi_{B21} = -0.639670778$	$\alpha_3 = 1.863548208$, $\psi_{B21} = -0.639165506$
48	0.18	0.99/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.873610909$, $\psi_{B21} = -0.639670778$	$\alpha_3 = 1.874071208$, $\psi_{B21} = -0.639165506$
49	0.18	0.60/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.860084729$, $\psi_{B21} = -0.636720778$	$\alpha_3 = 1.860545792$, $\psi_{B21} = -0.636215506$
50	0.18	0.80/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.863034729$, $\psi_{B21} = -0.636720778$	$\alpha_3 = 1.863495792$, $\psi_{B21} = -0.636215506$
51	0.18	0.99/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.873557729$, $\psi_{B21} = -0.636720778$	$\alpha_3 = 1.874018792$, $\psi_{B21} = -0.636215506$
52	0.18	0.60/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.859894456$, $\psi_{B21} = -0.626197778$	$\alpha_3 = 1.860357987$, $\psi_{B21} = -0.625692506$
53	0.18	0.80/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.862844456$, $\psi_{B21} = -0.626197778$	$\alpha_3 = 1.863307987$, $\psi_{B21} = -0.625692506$

Table 10: Values of the Key Parameters in the Population Model and Models Misspecified in Both the Marginal Mean & the Between-Level Covariance Structures (*Continued*)

		Severity		**Population Values**	
PM	**ICC**	(α_3/ψ_{B21})	**Misspecification**	**Balanced Data**	**Unbalanced Data**
54	0.18	0.99/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.873367456$, $\psi_{B21} = -0.626197778$	$\alpha_3 = 1.873830987$, $\psi_{B21} = -0.625692506$
55	0.25	0.60/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.770444567$, $\psi_{B21} = -0.852392799$	$\alpha_3 = 1.770664981$, $\psi_{B21} = -0.852126045$
56	0.25	0.80/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.773394567$, $\psi_{B21} = -0.852392799$	$\alpha_3 = 1.773614981$, $\psi_{B21} = -0.852126045$
57	0.25	0.99/0.60	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.783917567$, $\psi_{B21} = -0.852392799$	$\alpha_3 = 1.784137981$, $\psi_{B21} = -0.852126045$
58	0.25	0.60/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.770376330$, $\psi_{B21} = -0.849442799$	$\alpha_3 = 1.770597215$, $\psi_{B21} = -0.849176045$
59	0.25	0.80/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.773326330$, $\psi_{B21} = -0.849442799$	$\alpha_3 = 1.773547215$, $\psi_{B21} = -0.849176045$
60	0.25	0.99/0.80	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.783849330$, $\psi_{B21} = -0.849442799$	$\alpha_3 = 1.784070215$, $\psi_{B21} = -0.849176045$
61	0.25	0.60/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.770132171$, $\psi_{B21} = -0.838919799$	$\alpha_3 = 1.770354958$, $\psi_{B21} = -0.838653045$
62	0.25	0.80/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.773082171$, $\psi_{B21} = -0.838919799$	$\alpha_3 = 1.773304958$, $\psi_{B21} = -0.838653045$
63	0.25	0.99/0.99	$\alpha_3 = 4,\ \psi_{B21} = 0$	$\alpha_3 = 1.783605171$, $\psi_{B21} = -0.838919799$	$\alpha_3 = 1.783827958$, $\psi_{B21} = -0.838653045$

Table 11: Values of the Key Parameters in the Population Model and Models Misspecified in Both the Within-Level Covariance & Residual Structures

PM	ICC	Severity (ψ_{W32}/ζ_3)	Misspecification	Population Values: Balanced Data	Population Values: Unbalanced Data
64	0.10	0.60/0.60	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.323122304$, $\zeta_3 = 2.578909788$	$\psi_{W32} = 0.323718554$, $\zeta_3 = 2.578954179$
65	0.10	0.80/0.60	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.323122304$, $\zeta_3 = 3.168905788$	$\psi_{W32} = 0.323718554$, $\zeta_3 = 3.168950179$
66	0.10	0.99/0.60	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.323122304$, $\zeta_3 = 5.273619788$	$\psi_{W32} = 0.323718554$, $\zeta_3 = 5.273664179$
67	0.10	0.60/0.80	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.913118304$, $\zeta_3 = 2.399188180$	$\psi_{W32} = 0.913714554$, $\zeta_3 = 2.398953856$
68	0.10	0.80/0.80	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.913118304$, $\zeta_3 = 2.989184180$	$\psi_{W32} = 0.913714554$, $\zeta_3 = 2.988949856$
69	0.10	0.99/0.80	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.913118304$, $\zeta_3 = 5.093898180$	$\psi_{W32} = 0.913714554$, $\zeta_3 = 5.093663856$
70	0.10	0.60/0.99	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 3.017832304$, $\zeta_3 = 1.165720391$	$\psi_{W32} = 3.018428554$, $\zeta_3 = 1.164168084$
71	0.10	0.80/0.99	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 3.017832304$, $\zeta_3 = 1.755716391$	$\psi_{W32} = 3.018428554$, $\zeta_3 = 1.754164084$
72	0.10	0.99/0.99	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 3.017832304$, $\zeta_3 = 3.860430391$	$\psi_{W32} = 3.018428554$, $\zeta_3 = 3.858878084$
73	0.18	0.60/0.60	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.265931332$, $\zeta_3 = 2.530920117$	$\psi_{W32} = 0.266526018$, $\zeta_3 = 2.530941174$
74	0.18	0.80/0.60	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.265931332$, $\zeta_3 = 3.120916117$	$\psi_{W32} = 0.266526018$, $\zeta_3 = 3.120937174$
75	0.18	0.99/0.60	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.265931332$, $\zeta_3 = 5.225630117$	$\psi_{W32} = 0.266526018$, $\zeta_3 = 5.225651174$
76	0.18	0.60/0.80	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.855927332$, $\zeta_3 = 2.356116596$	$\psi_{W32} = 0.856522018$, $\zeta_3 = 2.355836795$
77	0.18	0.80/0.80	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.855927332$, $\zeta_3 = 2.946112596$	$\psi_{W32} = 0.856522018$, $\zeta_3 = 2.945832795$
78	0.18	0.99/0.80	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 0.855927332$, $\zeta_3 = 5.050826596$	$\psi_{W32} = 0.856522018$, $\zeta_3 = 5.050546795$
79	0.18	0.60/0.99	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 2.960641332$, $\zeta_3 = 1.143585100$	$\psi_{W32} = 2.961236018$, $\zeta_3 = 1.141839081$
80	0.18	0.80/0.99	$\psi_{W32} = 0,\ \zeta_3 = 12$	$\psi_{W32} = 2.960641332$, $\zeta_3 = 1.733581100$	$\psi_{W32} = 2.961236018$, $\zeta_3 = 1.731835081$

Table 11: Values of the Key Parameters in the Population Model and Models Misspecified in Both the Within-Level Covariance & Residual Structures (*Continued*)

PM	**ICC**	**Severity** (ψ_{W32}/ζ_3)	**Misspecification**	**Population Values** Balanced Data	**Population Values** Unbalanced Data
81	0.18	0.99/0.99	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 2.960641332$, $\zeta_3 = 3.838295100$	$\psi_{W32} = 2.961236018$, $\zeta_3 = 3.836549081$
82	0.25	0.60/0.60	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 0.175530976$, $\zeta_3 = 2.459712413$	$\psi_{W32} = 0.175891543$, $\zeta_3 = 2.459570746$
83	0.25	0.80/0.60	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 0.175530976$, $\zeta_3 = 3.049708413$	$\psi_{W32} = 0.175891543$, $\zeta_3 = 3.049566746$
84	0.25	0.99/0.60	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 0.175530976$, $\zeta_3 = 5.154422413$	$\psi_{W32} = 0.175891543$, $\zeta_3 = 5.154280746$
85	0.25	0.60/0.80	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 0.765526976$, $\zeta_3 = 2.294814514$	$\psi_{W32} = 0.765887543$, $\zeta_3 = 2.294391670$
86	0.25	0.80/0.80	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 0.765526976$, $\zeta_3 = 2.884810514$	$\psi_{W32} = 0.765887543$, $\zeta_3 = 2.884387670$
87	0.25	0.99/0.80	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 0.765526976$, $\zeta_3 = 4.989524514$	$\psi_{W32} = 0.765887543$, $\zeta_3 = 4.989101670$
88	0.25	0.60/0.99	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 2.870240976$, $\zeta_3 = 1.124668297$	$\psi_{W32} = 2.870601543$, $\zeta_3 = 1.122832265$
89	0.25	0.80/0.99	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 2.870240976$, $\zeta_3 = 1.714664297$	$\psi_{W32} = 2.870601543$, $\zeta_3 = 1.712828265$
90	0.25	0.99/0.99	$\psi_{W32} = 0$, $\zeta_3 = 12$	$\psi_{W32} = 2.870240976$, $\zeta_3 = 3.819378297$	$\psi_{W32} = 2.870601543$, $\zeta_3 = 3.817542265$

Other factors concerning the characteristics of longitudinal data may also influence the estimation and evaluation of linear-linear piecewise multilevel latent growth curve models. The most typical one is that the outcome measures are categorical (as in surveys), which violates the multivariate normality assumption for the model. Different estimation method is then required[18], which gives rise to different model fit indices. Another issue involves the missingness of observations across time, which is common for longitudinal data and which also affects the choice of the practical fit indices. There are still many other factors that are not included in the current study but are influential for the evaluation of linear-linear piecewise multilevel latent growth curve modeling (e.g., cluster/individual ratio, number of repeated measures, and number and location of the knot(s)). However, a decision is made to exclude all those factors so as to evaluate the model under an ideal condition.

Replications with convergence problems (i.e., the latent variable covariance matrix is not positive definite) are first deleted casewise from the final results. Then a series of analysis of variance tests are conducted to decide the significant factors that may influence the performance of the fit indices in detecting each type of misspecification, or misspecifications on the between-level, or those on the within-

---

[18] A recent development is to use robust standard errors (generally the Huber-White sandwich estimator (Huber, 1967; White; 1982)) and chi-squares for significant testing (usually the Satorra-Bentler (1994) and the Yuan-Bentler (1998) corrections) when violations of the assumptions of the asymptotic tests are suspected. With multilevel data, robust chi-squares and standard errors are assumed to offer some protection against unmodeled heterogeneity, which may result from misspecifying the group-level model, or by omitting a level (Hox et al., 2010).

level, or those across both levels. The dependent variable of each analysis of variance is the obtained value of a fit index. The independent variables include severity of misspecification, intraclass correlation coefficient, group-level sample size, individual-level sample size, and cluster balance if the type of misspecification is specified. When the type of misspecification is not known a prior, the factor “type of misspecification” is included as an independent variable in the analysis of variance. Furthermore, the analyses of variance include all possible interactions between these independent variables. Partial η^2 instead of η^2 is reported as the effect size for each factor, since the former is the variance explained by a given variable of the variance remaining after excluding variance explained by other predictors (i.e., its magnitude will not be reduced by including other factors into an experimental design, as is the case with η^2), which enables the comparison of the effect size of an identical manipulation across studies with different designs (Cohen, 1988). Subsequently, correlations among fit indices for detecting a particular type of model misspecification and correlations of a particular fit index in detecting various types of model misspecifications are calculated to check whether those fit indices work equally well in detecting the same type of misspecification and whether a fit index work consistently in detecting different types of misspecification.

The cutoff criteria proposed by Hu and Bentler (1999) are also evaluated under the circumstance of linear-linear piecewise multilevel latent growth curve models. On one hand, the Type I and Type II error rates in detecting different types of misspecifications when using the suggested cutoff values are reported. On the other hand, the powers for each fit index to reject misspecified models when controlling for

Type I error rates are calculated, with the hope to find out new set of cutoff values that could work across different types of misspecifications.

The linear-linear piecewise multilevel latent growth curves are generated and analyzed via *Mplus* 7 (Muthén & Muthén, 1998-2012); and model fit information is collected, calculated and compared via SAS 9.3 (SAS Institute, 2002-2010).

3.2 Simulation Design 2

Previous study indicates that all the fit indices are sensitive to misspecifications in the marginal mean structure, in the within-level covariance structure, in the within-level residual structure, in both the marginal mean and the between-level covariance structures, and in both the within-level covariance and residual structures. The only type of misspecification to which none of the fit indices are sensitive is the one associated with the between-level covariance structure (i.e., ψ_{B_21} in this case). Hence the purpose of this study is to investigate the influence of two estimation methods on the sensitivity of the chosen practical fit indices (i.e., *NFI*, *TLI*, *CFI*, *Mc*, *RMSEA*, and *SRMR*) in detecting misspecification in the between-level covariance structure of a linear-linear piecewise multilevel latent growth curve model.

Given the fact that cluster balance does not influence the performance of all the fit indices, it is excluded from this study. Thus only five factors are manipulated in this study, which are group number (i.e., N_g = 50, 100, and 200), group size (i.e., N_{gi} = 10, 20, and 30), intraclass correlation coefficient (i.e., *ICC* = 0.10, 0.18, and 0.25), severity level (i.e., 0.60, 0.80, and 0.99) and estimation method (i.e., simultaneous estimation and partially-saturated estimation methods). Since both the true model and misspecified model are evaluated via both estimation methods, altogether there are $N_g(3)\times N_{gi}(3)\times ICC(3)\times$Severity(3)×Estimation(2)×Model(2) = 324 (conditions) included in the study. 200 replications[19] are adopted because the data have been pre-selected and have really high convergence rates.

3.3 Real Data Analyses

Previous research on latent growth curve modeling often uses artificial parameter values that are rarely seen in longitudinal data (e.g., Leite & Stapleton, 2011; Wu & West, 2010), which consequently makes those studies solely theoretical. In order to reinforce the practicality of the current simulation, this study uses growth model and parameter estimates that are obtained from the Longitudinal Study of American Youth (1987-1994, and 2007). As is stated in its user guide, the Longitudinal Study of American Youth is a national longitudinal database that is

> designed to examine the development of: (1) student attitudes toward and achievement in science, (2) student attitudes toward and achievement in mathematics, and (3) student interest in and plans for a career in science, mathematics, or engineering, during middle school, high school, and the first four years post-high school, and to estimate the relative influence of parents, home, teachers, school, peers, media, and selected informal learning experiences on these developmental patterns (Miller, 2011, pp. 1).

There are two cohorts included in the database, with Cohort One consisting of a national sample of 2,829 tenth-grade students who were followed for seven years, ending four years after high school in 1994, and Cohort Two consisting of a national sample of 3,116 seventh-grade students who were also tracked for seven years,

[19] 200 replications are the number adopted by Fan & Sivo (2005, 2007).

ending one year after high school in 1994[20]. The data used for the current simulation are the aggregate science scores collected in each fall semester from students in Cohort Two, who are evaluated for six consecutive years (i.e., 1987-1993, from their 7th grade to their 12th grade) with items developed by the NAEP. Out of the need of score interpretation, this test has an IRT score outcome with a mean of 50 and a standard deviation of 10, and is imputed for missing observations using different weight designs[21]. The lattice and spaghetti plots of the data suggest that a piecewise linear function may well explain students' growth trajectories over time. Moreover, the spaghetti plots of the two types of scores[22] suggest that score imputation changes the growth shape of the data (see Figure 4). Thus only IRT scores for individuals with complete observations over the six years are used for this analysis, ending up with a sample of 567 students nested within 48 schools.

[20] Pure clustering of students within schools is not achieved in this case. However, this issue is ignored because this is not a study to investigate the influence of a particular school.

[21] Detailed weighting designs are listed in the user guide (pp. 29-30).

[22] Cases with missing observations were deleted for both IRT scores and imputed scores, which however is not necessary for multilevel growth curve modeling.

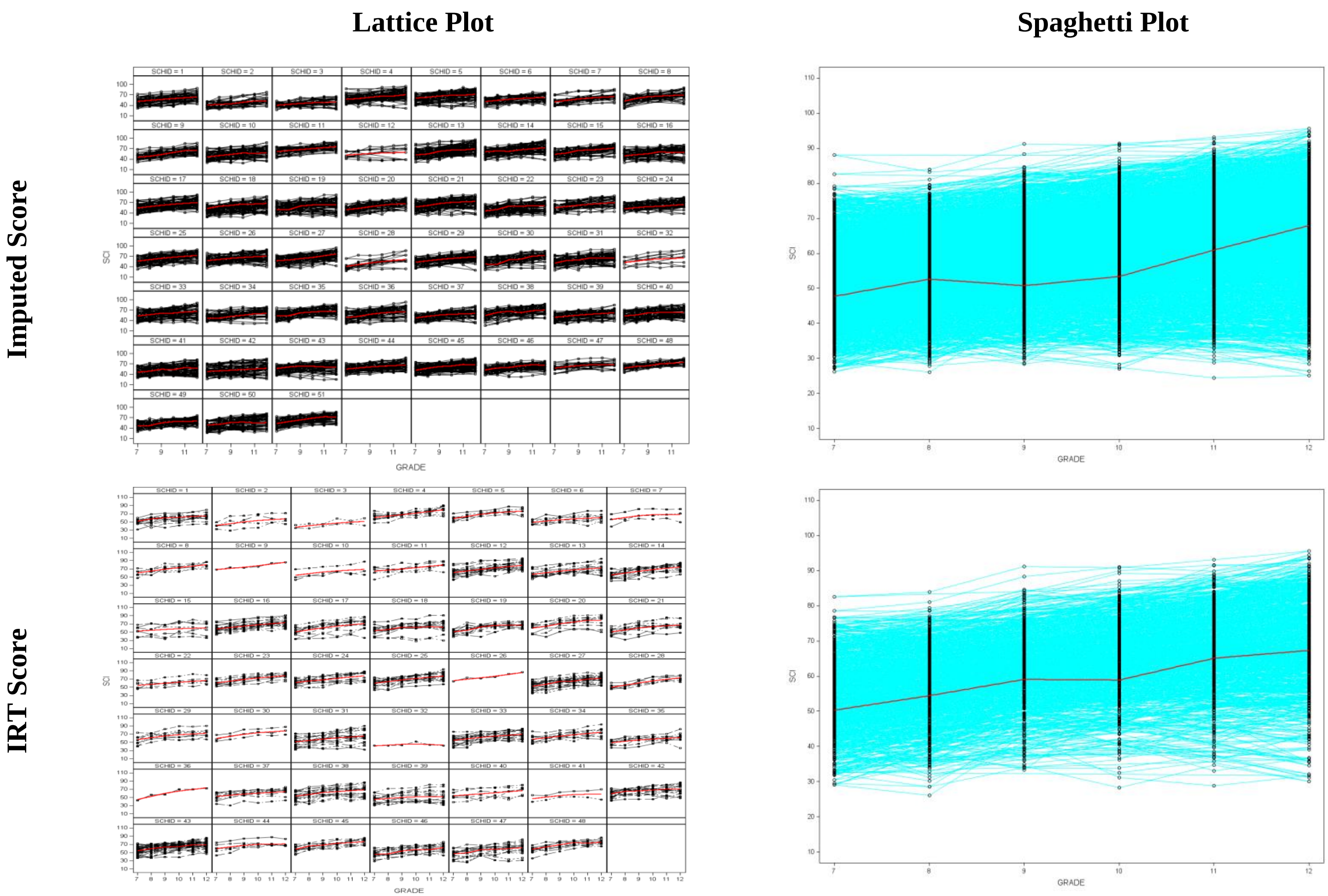


Figure 4: Lattice and spaghetti plots for LSAY science data with or without score imputation.

The following steps are adopted to obtain a linear-linear piecewise multilevel latent growth curve model from the LSAY data:

Step 1: Decide the approximate number and location of the knot(s).

Several methods are available to select the appropriate number and position of time knots for a multilevel growth curve model (see Howe, Tilling, Matijasevich, Petherick, Santos, Fairley, Wright, Santos, Barros, Martin, Krammer, Bogdanovich, Matush, Barros, & Lawlor, 2013). The most typical one is to use fractional polynomials to derive a smooth function for the curve, whose derivatives are then calculated to decide the number and position of the knots (e.g., Ben-Shlomo, McCarthy, Hughes, Tilling, Davies, & Smith, 2008; Howe, Tilling, Galobardes, Smith, Gunnell, & Lawlor, 2012). A reverse method is to start with a large number of knots, and gradually reduce the number until a smooth curve is achieved. Other options include placing knots at the centiles of the distribution of the time variable, or using stepwise regression to decide whether the linear slopes on either side of a knot point is statistically significantly different. A preliminary analysis of the IRT scores suggests that an exponential function may explain the LSAY data, based on which one knot is found to occur at the fourth time point. Then additive time coding and stepwise regressions are used to confirm that conclusion. In order to validate the coding method, a small simulation (i.e., based on a three-level linear-linear piecewise growth curve model with the knot being placed at the 4th time point) is conducted to check whether time coding method (i.e., additive versus piecewise) influences growth parameter estimates and model fit indices[23]. The results suggest that time coding method does not statistically influence either the fixed effects, the random effects at different levels[24], or the model fit indies, given the model is ‘correct’ (See Figures 5-9).

[23] The relationship of the mean and the variance-covariance structures under piecewise and additive time coding methods is illustrated in Appendix.

[24] Bias, standard error (SE), and root-mean-squared-error (RMSE) may be adopted to calibrate the parameter recovery for different time coding methods.

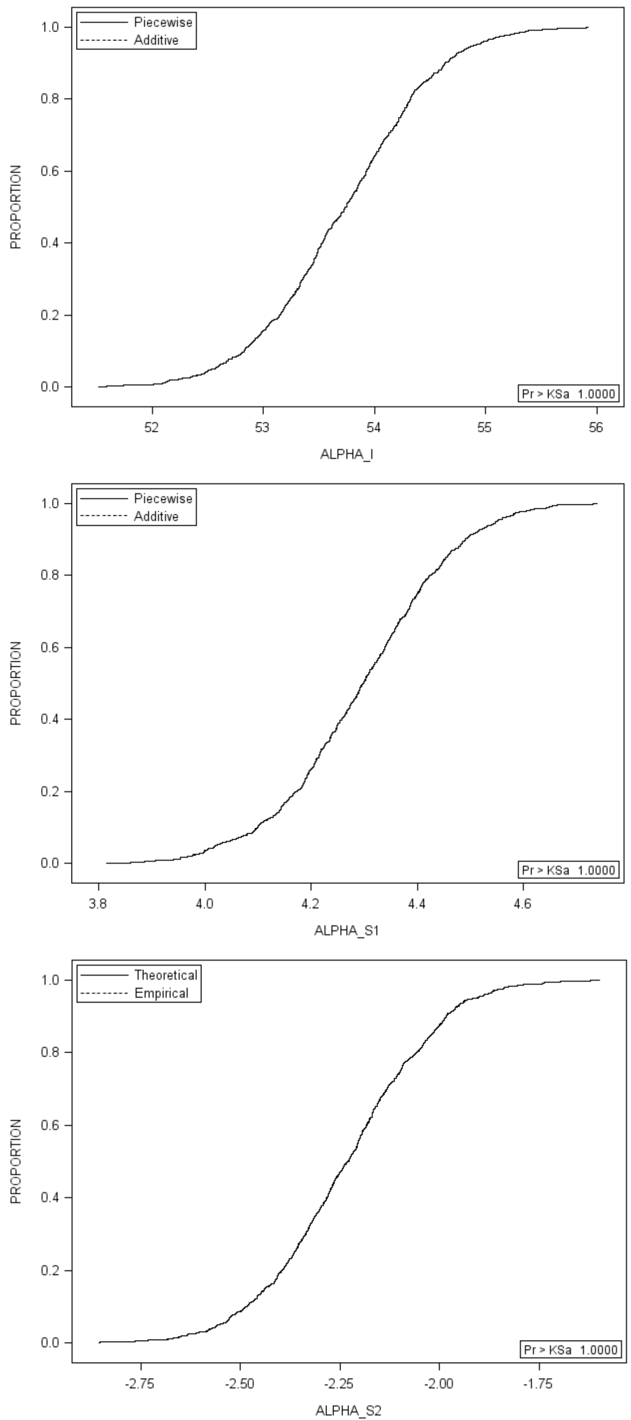


Figure 5: Empirical distribution function plots of the fixed effects produced via piecewise & additive time coding.

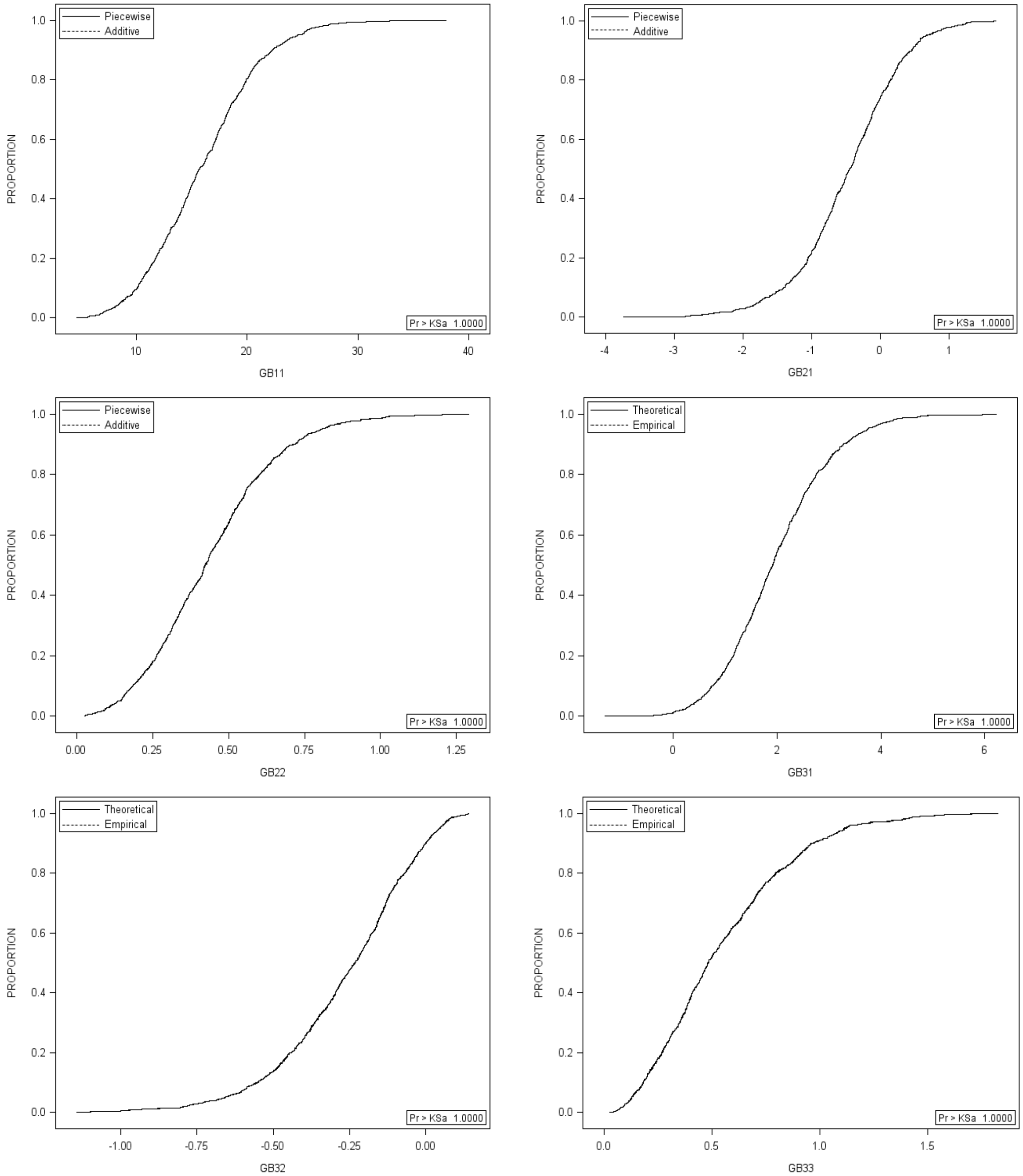


Figure 6: Empirical distribution function plots of the random between-level effects produced via piecewise & additive time coding.

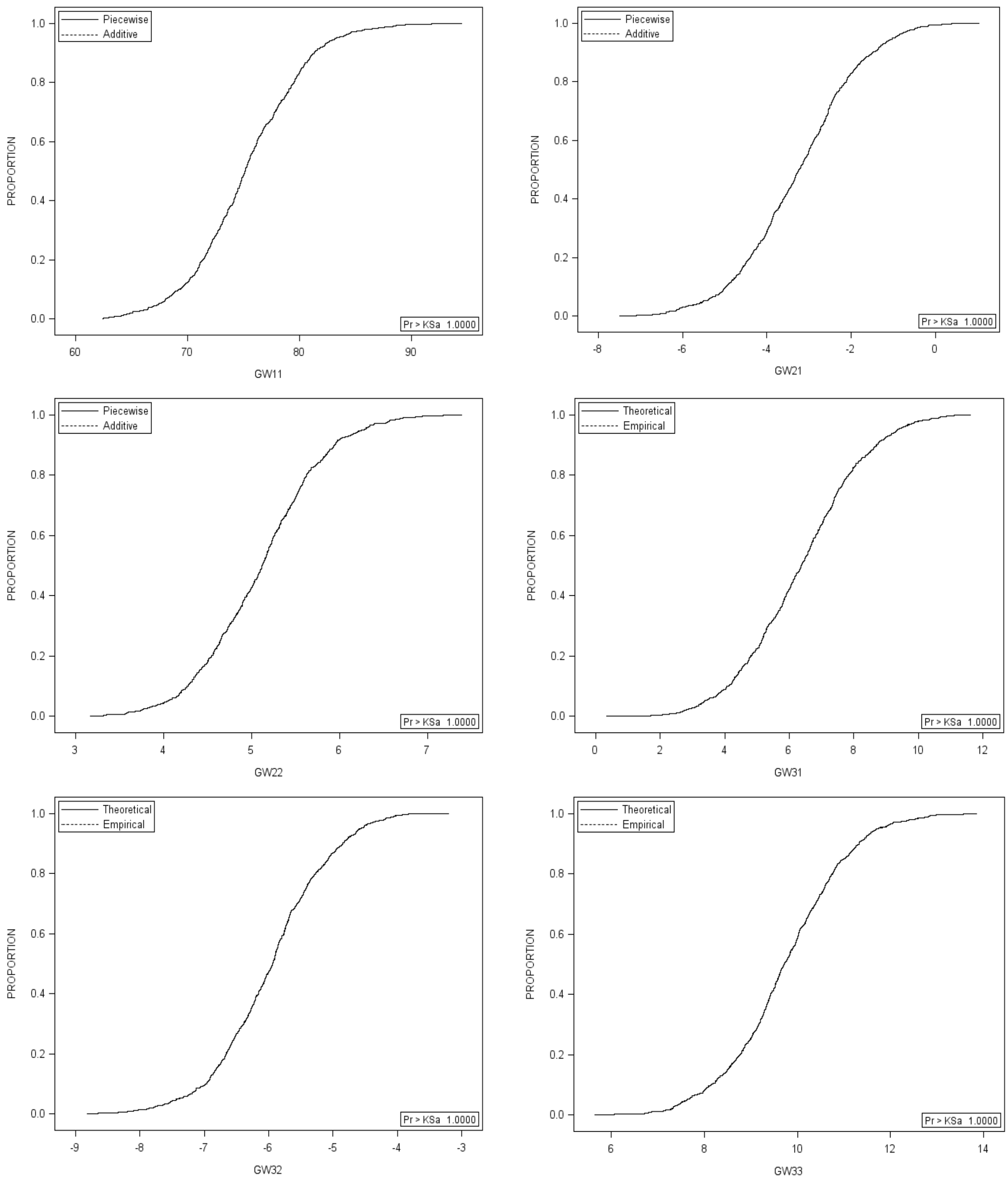


Figure 7: Empirical distribution function plots of the random within-level effects produced via piecewise & additive time coding.

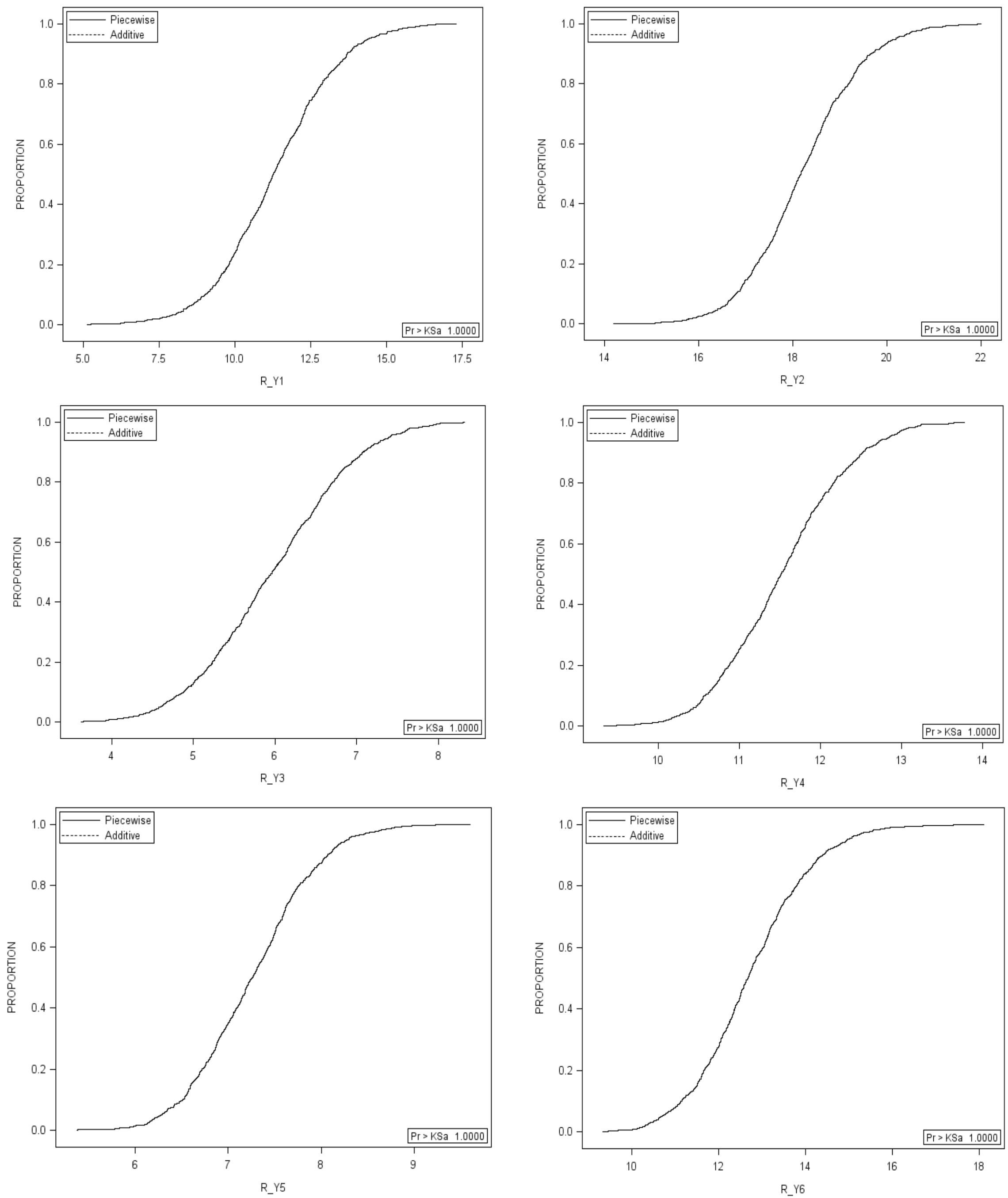


Figure 8: Empirical distribution function plots of the residuals produced via piecewise & additive time coding.

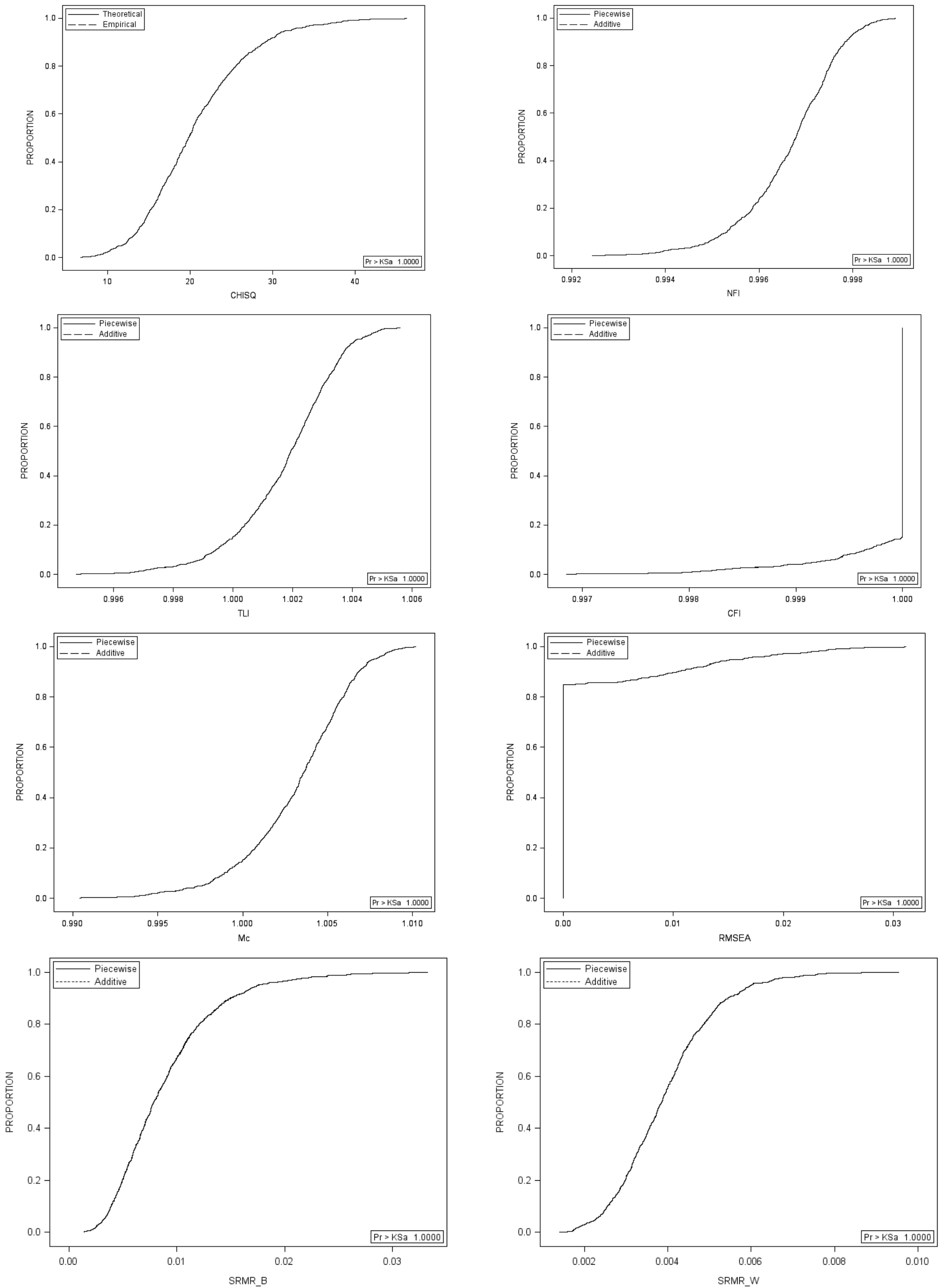


Figure 9: Empirical distribution function plots of the fit indices for models estimated via piecewise & additive time coding.

Step 2: Fit a linear-linear piecewise multilevel growth curve model to the LSAY data by using SAS Proc Mixed procedure (see Table 12).

Table 12: Multilevel Growth Analysis for LSAY Science Data Via SAS ML Estimation

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$$\mathbf{\Gamma} = \begin{pmatrix} 53.708 \\ 4.288 \\ 2.056 \end{pmatrix}$$

$$\mathbf{T}_B = \begin{pmatrix} 17.141 & & \\ -0.558 & 0.488 & \\ 1.555 & 0.192 & 0.486 \end{pmatrix}$$

$$\mathbf{T}_W = \begin{pmatrix} 75.144 & & \\ -3.211 & 5.120 & \\ 3.139 & -0.881 & 2.933 \end{pmatrix}$$

$$\mathbf{\Sigma}_r = \begin{pmatrix} 11.408 & & & & & \\ & 18.178 & & & & \\ & & 5.981 & & & \\ & & & 11.568 & & \\ & & & & 7.303 & \\ & & & & & 12.646 \end{pmatrix}$$

a. Fit a 3-level piecewise model with a loaded mean structure, with the location of knot being decided by the previous step;

b. Select a covariance structure for both the between- and the within-group random effects. A series of likelihood ratio tests using the maximum likelihood estimation[25] method were implemented to select the superior covariance structure for each level. Because the test statistic (i.e., the random effects) were at the boundary of the parameter space (i.e., 0), a mixture of χ^2 distributions, with each having a weight of 0.5[26] was adopted to test the hypotheses (Verbeke & Molenberghs, 2000).

c. Select a covariance structure for the residuals. Fourteen residual covariance structures (i.e., variance components, first-order autoregressive, heterogeneous first-order autoregressive, compound symmetry, heterogeneous compound symmetry, Toeplitz, heterogeneous Toeplitz, unstructured, banded main diagonal, spatial exponential, spatial Gaussian, spatial linear, spatial power, and spatial spherical) were tested by using the likelihood ratio tests and compared by using the information

---

[25] The maximum likelihood estimation method is used here to correspond to the estimation method that is integrated in Mplus. This method, however, produces biased covariance parameter estimates and deflated standard errors for the estimates of the fixed effects (West, Welch, & Galecki, 2007). In contrast, the restricted maximum likelihood estimation is usually suggested in multilevel growth curve modeling to address that issue.

[26] Different weight designs may be adopted based on the specific distribution of the real data. In addition, this sample size is not typically sufficiently large for a mixture analysis.

based fit indices such as AIC, BIC, and AICc to reflect the continuous heterogeneous distribution of the residuals and to satisfy the condition that not all of the covariance structures were nested within each other.

d. Conduct diagnostic analyses for the selected model, including the normality and homoscedasticity for the residuals, and the multivariate normality of the predicted values of the random effects (i.e, EBLUPs) at different levels (see Figures 10-13)[27]. Test results indicate that the skewness of all the EBLUPs of the random effects as well as that for the residuals roughly fall within the range of [-0.5, 0.5], suggesting that they are approximately symmetric (Bulmer, 1979). The kurtosis values, however, varies from -0.4 to 1.2, suggesting that they are either mildly platykurtic or leptokurtic. A problem reflected by the diagnostics is that the EBLUPs for the group-level linear slopes suggests a mixture of two groups of schools, which undermined the validity of current one-group statistical model. Even if the small group has minor influence on the parameter estimates, the fact that kurtosis could have considerable impact on significance tests and standard errors of parameter estimates (Finch et al., 1997; Mardia, Kent, & Bibby, 1979) hazards the growth parameter estimates that are obtained for the current linear-linear piecewise multilevel growth curve model. Considering the fact that the violation of the normality assumption is moderate and only a one-group model is

[27] Multiple univariate normality assumption is checked, which, however, is a necessary but not sufficient condition for multivariate normality (DeCarlo, 1997).

needed for this dissertation study, the statistical model is still retained and evaluated via *Mplus*.

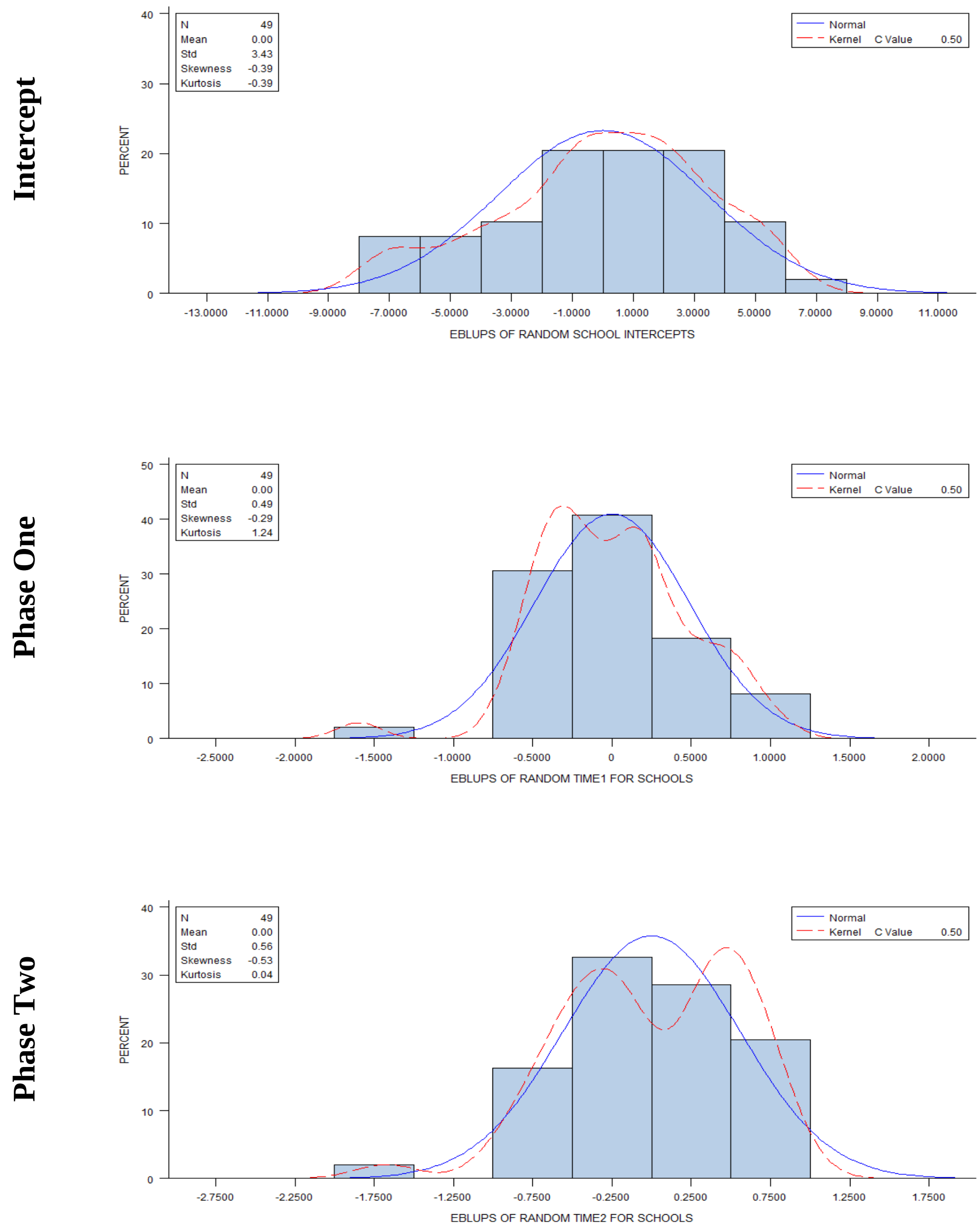


Figure 10: Histograms for the EPLUS of the school-level random effects for LSAY science data.

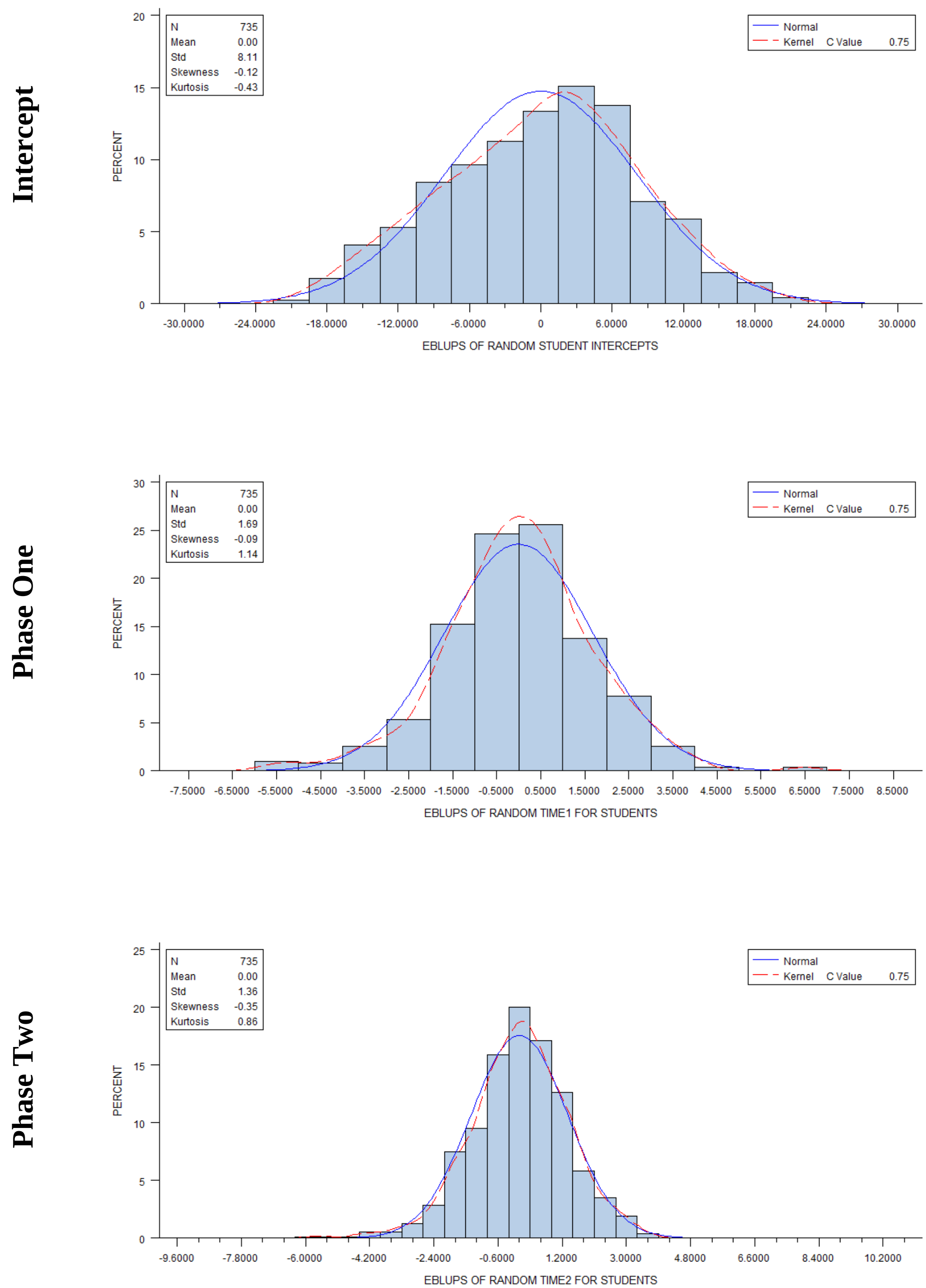


Figure 11: Histograms for the EPLUS of the student-level random effects for LSAY science data.

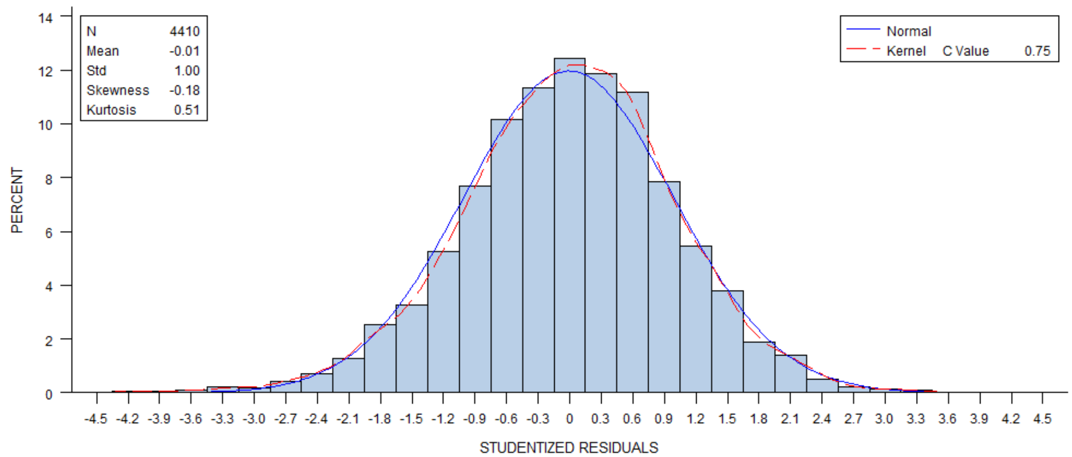


Figure 12: Histogram for the studentized residuals for LSAY science data.

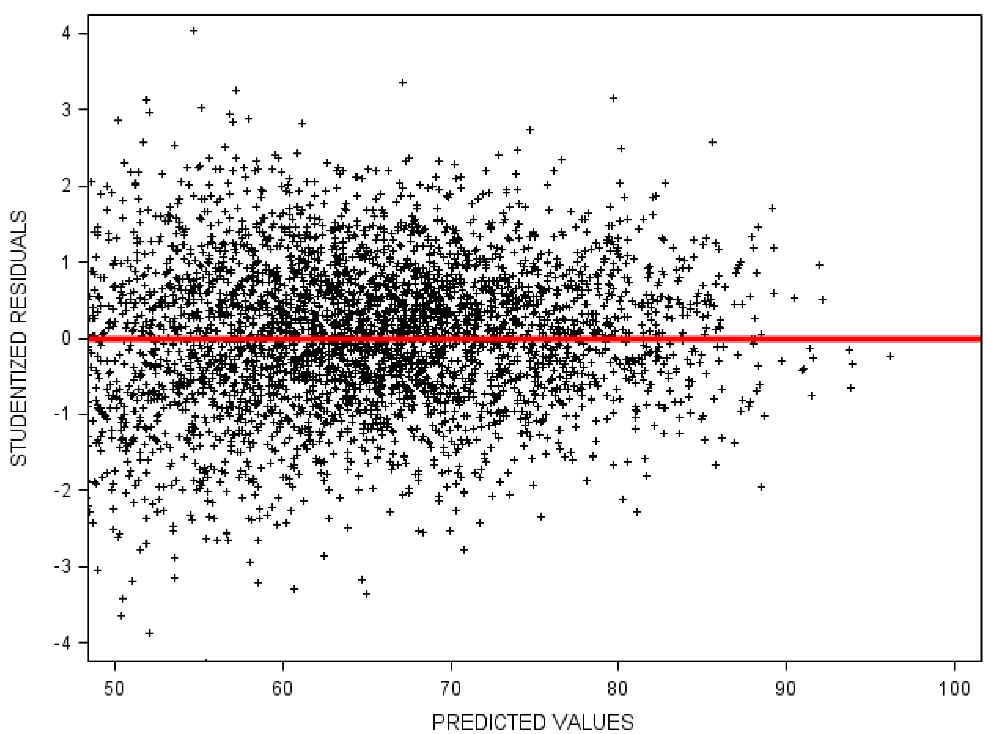


Figure 13: Homoscedasticity test for the multilevel growth model for LSAY science data.

Step 3: Fit the model obtained from Step 2 by using *Mplus* 7 (Muthén & Muthén, 1998-2012) so as to obtain the population growth parameters for the linear-linear piecewise multilevel latent growth curve model (see Table 13). This step could be conducted because (1) a multilevel growth curve model and a multilevel latent growth curve model converge on most occasions, with the latter being more precise than the former (Bauer, 2003; Curran, 2003; Raudenbush, 2001; Willett & Sayer, 1994), and (2) it is still complicated to use *Mplus* to figure out a linear piecewise multilevel growth curve model directly from a particular dataset (see Kohli & Harring, 2013; Kwok, Luo, & West, 2010). Because the multivariate normality assumption is moderately violated, both the traditional maximum likelihood estimation and the maximum likelihood estimation using a non-normality robust standard error approach (MLR) are applied to obtain the growth parameter values. As is reported by Hox and his research fellows (2010), these two estimation methods do not influence growth parameter estimates but their standard errors and the chi-square tests.

Table 13: Multilevel Latent Growth Analysis for LSAY Science Data via Mplus ML & MLR Estimation

ML Estimation	MLR Estimation
$\boldsymbol{\mu} = \begin{pmatrix} 53.710 \\ 4.289 \\ 2.056 \end{pmatrix}$	$\boldsymbol{\mu} = \begin{pmatrix} 53.710 \\ 4.289 \\ 2.056 \end{pmatrix}$
$\boldsymbol{\Psi}_B = \begin{pmatrix} 16.581 & & \\ -0.537 & 0.461 & \\ 1.513 & 0.191 & 0.467 \end{pmatrix}$	$\boldsymbol{\Psi}_B = \begin{pmatrix} 16.581 & & \\ -0.537 & 0.461 & \\ 1.513 & 0.191 & 0.467 \end{pmatrix}$
$\boldsymbol{\Psi}_W = \begin{pmatrix} 75.155 & & \\ -3.209 & 5.121 & \\ 3.139 & -0.880 & 2.933 \end{pmatrix}$	$\boldsymbol{\Psi}_W = \begin{pmatrix} 75.154 & & \\ -3.209 & 5.121 & \\ 3.139 & -0.880 & 2.933 \end{pmatrix}$
$\boldsymbol{\Theta} = \begin{pmatrix} 11.409 & & & & & \\ & 18.178 & & & & \\ & & 5.985 & & & \\ & & & 11.565 & & \\ & & & & 7.302 & \\ & & & & & 12.694 \end{pmatrix}$	$\boldsymbol{\Theta} = \begin{pmatrix} 11.409 & & & & & \\ & 18.178 & & & & \\ & & 5.985 & & & \\ & & & 11.565 & & \\ & & & & 7.302 & \\ & & & & & 12.694 \end{pmatrix}$

Chapter 4: Results

4.1 Results of Study 1

4.1.1 Descriptive statistics

Model evaluation results are collected into a dataset, with replications that does not converge being deleted casewise[28] and excluded from further analysis. As is listed in Table 14, most conditions converge completely, and the non-convergence rate is low for those which do not converge completely[29] (i.e., less than 1%).

[28] "MITERATIONS = 100000" was set so as to reach maximum convergence rate.

[29] Population values used in the simulation study were preselected, meaning that the whole set of values were deleted if the convergence rate was less than 90%. Instead, another set of values for all the parameters were tried again until the criterion of 90% convergence rate was achieved. This 90% convergence rate criterion was decided arbitrarily by the author.

Table 14: Summary of the Convergence Rates for the True and Misspecified Models Failing to Achieve Complete Convergence (i.e., 100% Convergence)

					Convergence Rates (%)	
Model / Misspecification	**ICC**	**Severity**	**N_g**	**N_{gi}**	**Balanced Data**	**Unbalanced Data**
True	0.10	0.60	50	20	/	99.90
True	0.10	0.60	100	20	/	99.90
True	0.10	0.80	50	20	/	99.90
True	0.10	0.80	100	20	/	99.90
True	0.10	0.99	50	20	/	99.90
True	0.10	0.99	100	20	/	99.90
True	0.18	0.80	50	10	/	99.80
ψ_{B21}	0.10	0.99	50	20	/	99.90
ψ_{B21}	0.25	0.60	50	10	/	99.90
ψ_{B21}	0.25	0.80	50	10	/	99.90
ψ_{W32}	0.10	0.80	50	10	99.80	/
ζ_3	0.10	0.60	100	10	99.90	/
ζ_3	0.10	0.80	50	10	99.90	99.90
ζ_3	0.10	0.80	50	20	/	99.90
ζ_3	0.25	0.99	50	10	99.90	/
α_3 & ψ_{B21}	0.18	0.60/0.99	50	10	99.90	/
α_3 & ψ_{B21}	0.18	0.80/0.60	50	10	99.90	/
α_3 & ψ_{B21}	0.18	0.80/0.80	50	10	99.90	/
α_3 & ψ_{B21}	0.18	0.99/0.99	50	10	99.90	/
α_3 & ψ_{B21}	0.25	0.60/0.80	50	10	99.90	/
α_3 & ψ_{B21}	0.25	0.99/0.80	50	10	99.90	/
ψ_{W32} & ζ_3	0.10	0.60/0.80	50	10	/	99.90

Note: The convergence rates were calculated based on 1000 replications for each condition.

The descriptive statistics of the performance of the fit indices across true and misspecified models are listed in Table 15. It is expected to see an ideal mean value with a small standard deviation for all the fit indices if a model is true; on the contrary, it is expected to see mean values departure largely from the ideal ones if a model is misspecified. The simulation results suggest that all the fit indices perform well with the true model. The means of *NFI*, *TLI*, *CFI*, and *Mc* are close to one and the means of *RMSEA* and *SRMR* are close to zero. In addition to the mean values of the practical fit indices, all of them have tiny standard deviations (i.e., smaller than 0.01) across different types of misspecifications, implying that all of them perform quite stably in detecting misspecifications in linear-linear piecewise multilevel latent growth curve models.

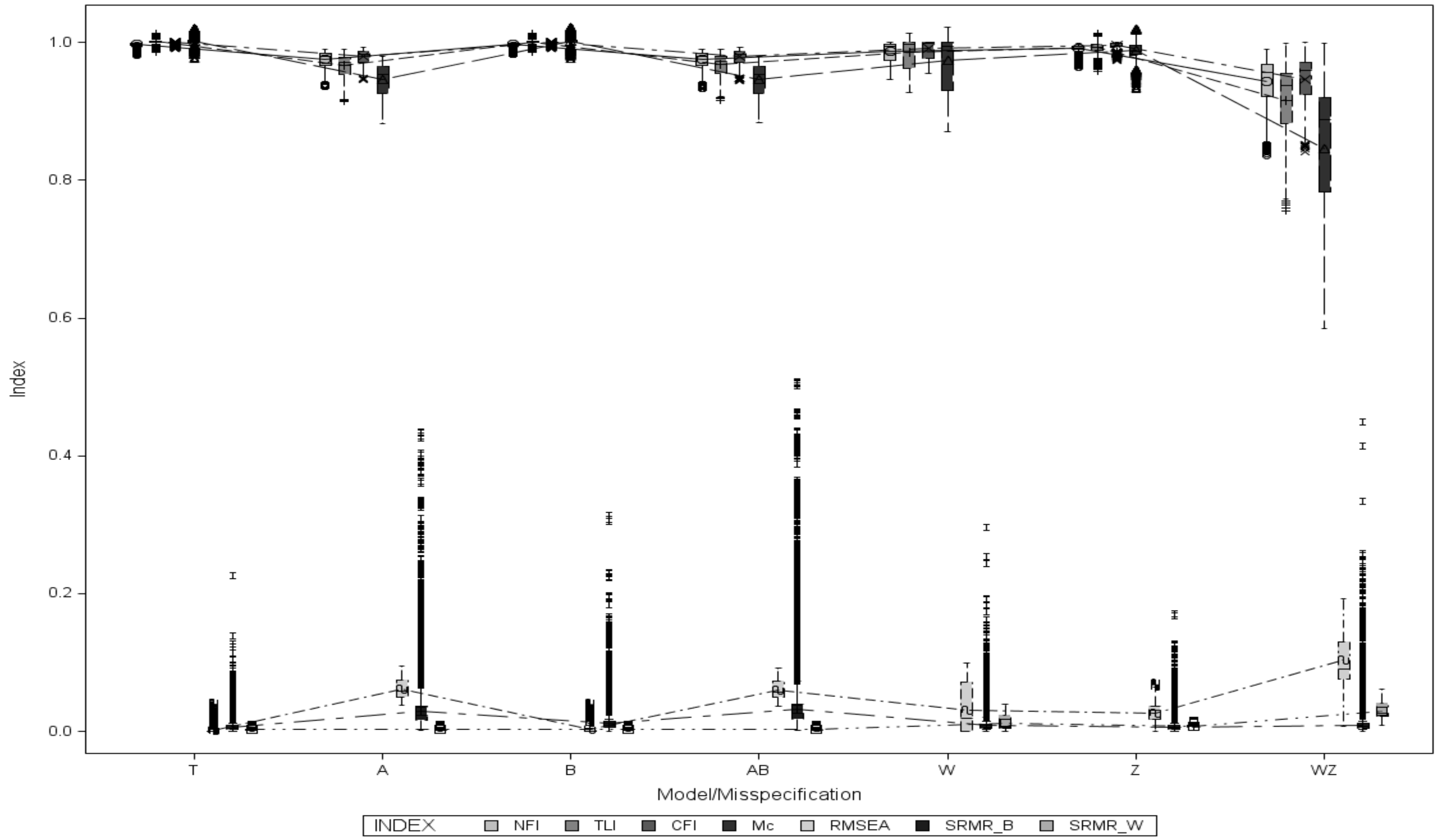
Table 15: Descriptive Statistics of the Fit Indices Across True and Misspecified Models

	True (N = 161992)		α_3 (N = 161994)		ψ_{B21} (N = 161985)		α_3 & ψ_{B21} (N = 485967)		ψ_{W32} (N = 161990)		ζ_3 (N = 161989)		ψ_{W32} & ζ_3 (N = 485638)	
Index	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
NFI	0.9972	0.0023	0.9762	0.0100	0.9970	0.0023	0.9760	0.0100	0.9879	0.0119	0.9921	0.0037	0.9429	0.0344
TLI	1.0013	0.0021	0.9674	0.0133	1.0009	0.0021	0.9685	0.0128	0.9859	0.0192	0.9930	0.0051	0.9162	0.0538
CFI	0.9999	0.0004	0.9797	0.0083	0.9998	0.0005	0.9797	0.0082	0.9909	0.0117	0.9956	0.0030	0.9460	0.0347
Mc	1.0021	0.0034	0.9471	0.0203	1.0016	0.0035	0.9471	0.0202	0.9739	0.0357	0.9881	0.0088	0.8464	0.0992
RMSEA	0.0018	0.0043	0.0612	0.0120	0.0028	0.0051	0.0602	0.0117	0.0314	0.0318	0.0268	0.0125	0.1030	0.0372
SRMR[30]	0.0067	0.0058	0.0288	0.0189	0.0118	0.0099	0.0322	0.0246	0.0119	0.0094	0.0076	0.0021	0.0300	0.0108

Note: N here refers to the number of cases that converged.

[30] The mean and standard deviation for the true model are based on SRMR_B. The mean and standard deviation for SRMR_W are 0.0026 and 0.0012 respectively.

When a model is misspecified, however, the fit indices respond differentially to different types of misspecifications. As is shown in Table 15 and Figure 14, all fit indices follow the same pattern when they evaluate linear-linear piecewise multilevel latent growth curve models: (1) they are most sensitive to the misspecifications in both the within-level covariance and the residual structures, followed by the misspecification involving either the marginal mean or both the marginal mean and the between-level covariance structures, and then by misspecification in the within-level covariance structure, and finally by the misspecification in the within-level residual structure. None of them are obviously sensitive to the misspecification in the between-level covariance structure. In addition, all of them have similar sensitivity to misspecifications in either the marginal mean structure or both the marginal mean and the between-level covariance structures. (2) The standard deviations of the fit indices are largest when models are misspecified in both the within-level covariance and the residual structures, which are followed by cases when models are misspecified in the within-level covariance structure, in the marginal mean structure, in both the marginal mean and the between-level covariance structures, in the within-level residual structure, and finally in the between-level covariance structure. (3) The results that the distributions of the fit indices evaluating models misspecified in the between-level covariance structure almost coincide with the distributions of the fit indices for the true model suggest that none of the fit indices is able to detect the misspecification in the between-level covariance structure. When looking across the fit indices whose desirable direction is “large” (i.e., *NFI*, *TLI*, *CFI*, and *Mc*), *Mc* is found to deviate most severely from its true value across all types of model misspecifications; when comparing the fit indices whose desirable direction is “small”, *RMSEA* is found to deviate more severely than *SRMR* from its true value across all types of model misspecifications.



Notes: T = True model; A = Model misspecified in the marginal mean (α_3) structure; B = Model misspecified in the between-level covariance (ψ_{B21}) structure; W = Model misspecified in the within-level covariance (ψ_{W32}) structure; Z = Model misspecified in the residual (ζ_3) structure; AB = Model misspecified in both the marginal mean (α_3) and the between-level covariance (ψ_{B21}) structures; and WZ = Model misspecified in both the within-level covariance (ψ_{W32}) and the residual (ζ_3) structures.

Figure 14: Distributions of the fit indices across different types of model/misspecifications.

4.1.2 ANOVA analyses

A series of analyses of variance is conducted to find out significant factors that influence the performance of the practical fit indices in detecting each of the six types of model misspecifications, or misspecifications on the between-level in general, or misspecifications on the within-level in general, or misspecifications across both within and between levels.

Since the purpose of the current study is to investigate the sensitivity of the practical fit indices in detecting misspecifications in linear-linear piecewise multilevel latent growth curve models, large effect sizes resulting from types of misspecification and severity of misspecification are desirable if the type of misspecification is unknown a priori. In the case that the type of misspecification is fixed, severity of misfit is the only factor that is expected to significantly influence the performance of the practical fit indices. In contrast, other factors including sample size, intraclass correlation coefficient, and cluster balance are not expected to influence the practical fit indices because effects from these factors mean that the practical fit indices do not perform stably across data conditions, which consequently makes it difficult to make consistent decisions about model fit based on these fit indices.

Cohen's rule of thumb[31] (1988) (i.e., 0.01 = small, 0.06 = medium, and 0.14 = large) is used to evaluate the effect size associated with each factor, and those whose

[31] Partial η^2 in factorial ANOVA arguably more closely approximates what η^2 would have been explained for a factor had it been a one-way ANOVA, thus it is appropriate to apply Cohen's rule of thumb here because it is a one-way ANOVA that gives rise to Cohen's rules of thumb.

partial η^2 values are no smaller than 0.06 are reported. This cutoff value is selected to ensure that the selected factors are significant both statistically and practically.

When the model is true (see Table 16), *NFI*, *TLI*, *Mc*, and *SRMR* are all largely affected by the group-level and individual-level sample sizes. Moreover, *NFI* is also moderately influenced by the interactional effect between group-level and individual-level sample sizes, which makes it the index that is most severely influenced by sample size across all the selected practical fit indices. *CFI* and *RMSEA*, however, are not affected by either factor, making them superior to other practical fit indices. In terms of the respective effect of group and individual level sample sizes, the group-level sample size is found to have a larger effect than the individual-level sample size on *NFI*, *TLI*, and *Mc*, indicating that independent observations provided more information in the calculation of these fit indices. Compared with other fit indices, *SRMR* is the only index that is sensitive to the intraclass correlation coefficient when detecting misspecifications on the between level.

Table 16: Effect Sizes for the Fit Indices with the True Model

	Factors			
Index	N_g	N_{gi}	$N_g \times N_{gi}$	***ICC***
NFI	0.3904	0.3099	0.0786	—
TLI	0.1429	0.1000	—	—
CFI	—	—	—	—
Mc	0.1442	0.0989	—	—
RMSEA	—	—	—	—
SRMR_B	0.1129	0.1607	—	0.1973
SRMR_W	0.3286	0.2763	—	—

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When the model is misspecified and the position of the misspecifications is unknown (see Table 17), all fit indices are found to be largely sensitive to types of misspecification, severity of misspecification, and the interaction between these two factors. Again, *SRMR* is the only index that is sensitive to the intraclass correlation coefficient when model misspecifications involving the between level of linear-linear multilevel latent growth curve models.

Table 17: Effect Sizes for the Fit Indices with Misspecified Models

	Factors			
Index	**Misspecification**	**Severity**	**Misspecification×Severity**	***ICC***
NFI	0.2363	0.2211	0.2326	—
TLI	0.2459	0.2258	0.2387	—
CFI	0.2437	0.2271	0.2385	—
Mc	0.2591	0.2277	0.2407	—
RMSEA	0.2696	0.1299	0.1801	—
SRMR_B	0.2929	—	—	0.1084
SRMR_W	0.6308	0.0979	0.1298	—

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When model misspecifications are confined to the between-level (see Table 18), *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are found to be largely influenced by types of misspecification and individual-level sample size. Moreover, the influence of types of misspecification is larger than that from the individual-level sample size for *TLI*, *CFI*, *Mc*, and *RMSEA*. *NFI*, on the contrary, is most severely affected by sample size when compared with other practical fit indices. In contrast to all other fit indices, *SRMR* is only sensitive to the intraclass correlation coefficient, making it a less desirable fit index in evaluating linear-linear piecewise multilevel latent growth curve models.

Table 18: Effect Sizes for the Fit Indices with Models Misspecified on the Between Level

	Factors		
Index	**Misspecification**	N_{gi}	***ICC***
NFI	0.2875	0.3139	—
TLI	0.3676	0.2274	—
CFI	0.3391	0.2454	—
Mc	0.3682	0.2154	—
RMSEA	0.5262	0.0964	—
SRMR_B	—	—	0.1645

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When model misspecifications are constrained to the within level (see Table 19), all the fit indices are largely influenced by the severity of misspecifications. About 55% of the between subjects variance in the severity level plus error for *NFI*, *TLI*, *CFI*, and *Mc* and about 37% of the variance for that factor plus error for *RMSEA* and *SRMR* are attributable to the severity levels of misspecifications. In addition, the interaction between types of misspecifications and severity levels of misspecifications also contribute about 7% to the variability of *RMSEA* and *SRMR*.

Table 19: Effect Sizes for the Fit Indices with Models Misspecified on the Within Level

	Factors	
Index	**Severity**	**Misspecification×Severity**
NFI	0.5526	—
TLI	0.5566	—
CFI	0.5573	—
Mc	0.5610	—
RMSEA	0.3629	0.0706
SRMR_W	0.3903	0.0637

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When the misspecification is bounded to the marginal mean structure (see Table 20), group size is witnessed to have substantial effect on the variability of *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA*, with their partial η^2 values lingering around 0.9. In contrast, *SRMR* is not practically significantly influenced by group size. Instead, it is sensitive to the intraclass correlation coefficient, as is in the case when the misspecification involves the between level of the model and the position of the misspecification is unknown.

Table 20: Effect Sizes for the Fit Indices with Models Misspecified in the Marginal Mean Structure

	Factors	
Index	N_{gi}	*ICC*
NFI	0.8959	—
TLI	0.8907	—
CFI	0.8907	—
Mc	0.9170	—
RMSEA	0.9241	—
SRMR_B	—	0.3589

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

Table 21: Effect Sizes for the Fit Indices with Models Misspecified in the Between-Level Covariance Structure

	Factors			
Index	N_g	N_{gi}	$N_g \times N_{gi}$	*ICC*
NFI	0.3852	0.3151	0.0767	—
TLI	0.1357	0.0854	—	—
CFI	—	—	—	—
Mc	0.1369	0.0848	—	—
RMSEA	—	—	—	—
SRMR_B	—	—	—	0.2705

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When the misspecification involves only the between-level covariance structure (see Table 21), *TLI* and *Mc* are found to be greatly influenced by the group-level sample size and moderately influenced by the individual-level sample size. *NFI* again becomes the index that is most sensitive to sample size – it is primarily influenced by both individual and group level sample sizes. In addition, it is moderately influenced by the interaction between the two types of sample sizes. *CFI*, *RMSEA* and *SRMR*, however, are not practically significantly sensitive to either type of sample size when the misspecification occurs with the between-level covariance structure. *SRMR* behaves differentially from other practical fit indices in that it is largely influenced by the intraclass correlation coefficient.

When the misspecifications involve both the marginal mean and the between-level covariance structures (see Table 22), group size is seen to principally and significantly affect the variability of *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA*. On the contrary, *SRMR* is largely affected by the intraclass correlation coefficient, as is in the cases when models are misspecified in either the marginal mean structure or in the between-level covariance structure.

Table 22: Effect Sizes for the Fit Indices with Models Misspecified in Both the Marginal Mean and the Between-Level Covariance Structures

	Factors	
Index	N_{gi}	***ICC***
NFI	0.8937	—
TLI	0.8892	—
CFI	0.8892	—
Mc	0.9158	—
RMSEA	0.9227	—
SRMR_B	—	0.2135

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When the misspecification is found in the within-level covariance structure (see Table 23), all fit indices are decisively affected by the severity of misspecification, with their partial η^2 values vibrating around 0.95. This result suggests that all practical fit indices perform well in detecting misspecifications in the within-level covariance structure of linear-linear piecewise multilevel latent growth curve models.

Table 23: Effect Sizes for the Fit Indices with Models Misspecified in the Within-Level Covariance Structure

	Factor
Index	**Severity**
NFI	0.9384
TLI	0.9593
CFI	0.9635
Mc	0.9624
RMSEA	0.9558
SRMR_W	0.9646

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When a model is misspecified in the within-level residual structure (see Table 24), *TLI*, *CFI*, *Mc*, *RMSEA*, and *SRMR* are all found to perform well in that they are principally and significantly affected by the severity levels of the misspecification. Compared with other fit indices, *NFI* is a less desirable index since it is also largely and significantly affected by both group level and individual level sample sizes.

Table 24: Effect Sizes for the Fit Indices with Models Misspecified in the Within-Level Residual Structure

	Factors		
Index	**Severity**	N_g	N_{gi}
NFI	0.4283	0.1563	0.1262
TLI	0.6042	—	—
CFI	0.6223	—	—
Mc	0.6090	—	—
RMSEA	0.6259	—	—
SRMR_W	0.6418	—	—

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

When a model is misspecified in both the within-level covariance and residual structures (see Table 25), all fit indices are decisively and significantly influenced by the severity levels of both types of model misspecifications. Between the two types of model misspecifications, all fit indices are more sensitive to the severity level of the misspecification in the within-level covariance structure than that in the within-level residual structure.

Table 25: Effect Sizes for the Fit Indices with Models Misspecified in Both the Within-Level Covariance and Residual Structures

	Factors	
Index	**Severity_ψ_{32}**	**Severity_ζ_3**
NFI	0.8367	0.1254
TLI	0.8390	0.1257
CFI	0.8390	0.1257
Mc	0.8477	0.1226
RMSEA	0.8368	0.1449
SRMR_W	0.9118	0.0614

Note: An em dash indicates that the partial η^2 values are smaller than 0.06.

In sum, different types of model misspecifications, severity levels of a misspecification, as well as their interactions have large and significant effects on all fit indices when the position of a misspecification is unknown. When the misspecification occurs only on the between-level of the model, however, the variability of all fit indices is greatly influenced not only by types of misspecification but also by group size, which consequently undermines the practicality of the fit indices. When the misspecifications involve only the between-level model and the type of misspecification is fixed, *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are principally and significantly affected by group size, with *NFI* being the one that is most sensitive to sample size. *SRMR*, however, behaves differentially to other fit indices when the misspecifications are bounded to the between-level of the model – it is primarily and significantly influenced by the intraclass correlation coefficient. In other words, none of the practical fit indices is effective in detecting the misspecification in the between-level structure of linear-linear piecewise multilevel latent growth curve models because they are largely and significantly influenced either by sample sizes (i.e., *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA*) or by the intraclass correlation coefficient (i.e., *SRMR*).

When the misspecifications are unknown but are confined to the within-level of the model, all fit indices are greatly and significantly influenced by the severity level of the misspecifications. *RMSEA* and *SRMR* are additionally moderately sensitive to the interactions between types of misspecifications and severity levels of misspecifications. When the misspecifications involve both the within-level covariance and the within-level residual structures, all fit indices are primarily influenced by the severity level of the misspecification in the within-level covariance structure, followed by the severity levels of the misspecification in the within-level residual structure, both of which are large and statistically and practically significant. If the misspecification involves only the within-level covariance structure, all fit indices are decisively influenced by the severity level of the misspecification. If the misspecification involves solely the within-level residual structure, however, the variability of *NFI* is also largely and significantly influenced by individual and group level sample sizes. In other words, all fit indices perform well in detecting misspecifications in the within-level structure of linear-linear piecewise multilevel latent growth curve models. *NFI*, however, is less trustworthy among all fit indices because it is also sensitive to different levels of sample sizes.

Although it is not endorsed numerically by the ANOVA results, the distributions of the fit indices suggest that they are more sensitive to the misspecification in the marginal mean structure when both the marginal mean and the between-level covariance structures are misspecified. This is because the calculation of the covariance structure is based on the discrepancy between the observed and the estimated marginal means whereas the marginal means are far less sensitive to the specification of the covariance structure. To be more specific, constraining the marginal mean structure means to fit a linear function to the data generated from a linear-linear piecewise multilevel latent growth curve model. However, the marginal mean for linear models is asymptotically independent of the covariance structure for longitudinal data when the data are complete and balanced on time (Verbeke & lesaffre, 1997; Yuan & Bentler, 2004).

4.1.3 Graphical analyses

ANOVA analyses reveal that all fit indices are sensitive to sample size if the linear-linear piecewise multilevel latent growth curve model is misspecified on the between level. In addition, all of them are sensitive to the severity levels of misspecifications if the model is misspecified on the within level. Furthermore, *SRMR* is sensitive to the intraclass correlation coefficient when the misspecification involves the between-level structure of the model. All those effects are further analyzed and demonstrated in figures.

Effect of severity level. As is shown in Figures 15-20, *NFI*, *TLI*, *CFI*, and *Mc* share the same pattern when corresponding to different severity levels of the misspecifications occurring on the within-level of the model. When the model is misspecified in the within-level covariance structure, the means of *NFI*, *TLI*, *CFI*, and *Mc* decrease mildly when the severity level increase from 0.60 to 0.80, but sharply from 0.80 to 0.99, meaning that they have more power to reject misspecified models with the increment of the severity level in the misspecified within-level covariance structure. An opposite trend is observed if the misspecification involves the within-level residual structure, meaning that they are more likely to commit Type II errors with the increment of severity level in the misspecified within-level residual structure. When the misspecifications happen in both the within-level covariance and the residual structures, the average *NFI*, *TLI*, *CFI*, and *Mc* are noticed to decrease mildly within each severity level of the misspecified within-level residual structure when the severity level of the misspecified within-level covariance structure increase from 0.60 to 0.80. Moreover, these four fit indices increase sharply when the severity level of the misspecified within-level covariance structure rises from 0.80 to 0.99 within each severity level of the misspecified within-level residual structure. Within each level of the misspecified within-level covariance structure, the average *NFI*, *TLI*, *CFI*, and *Mc* are found to increase with the increment of the severity levels in the within-level residual structure. An opposite trend is found with *RMSEA* and *SRMR* in detecting misspecifications on the within level structure of the linear-linear piecewise multilevel latent growth curve model.

In addition, the spread of all fit indices increases if the model is misspecified in both the within-level covariance and residuals structures. However, they remain comparatively stable if only one type of misspecification is involved on the within-level structure of the model.

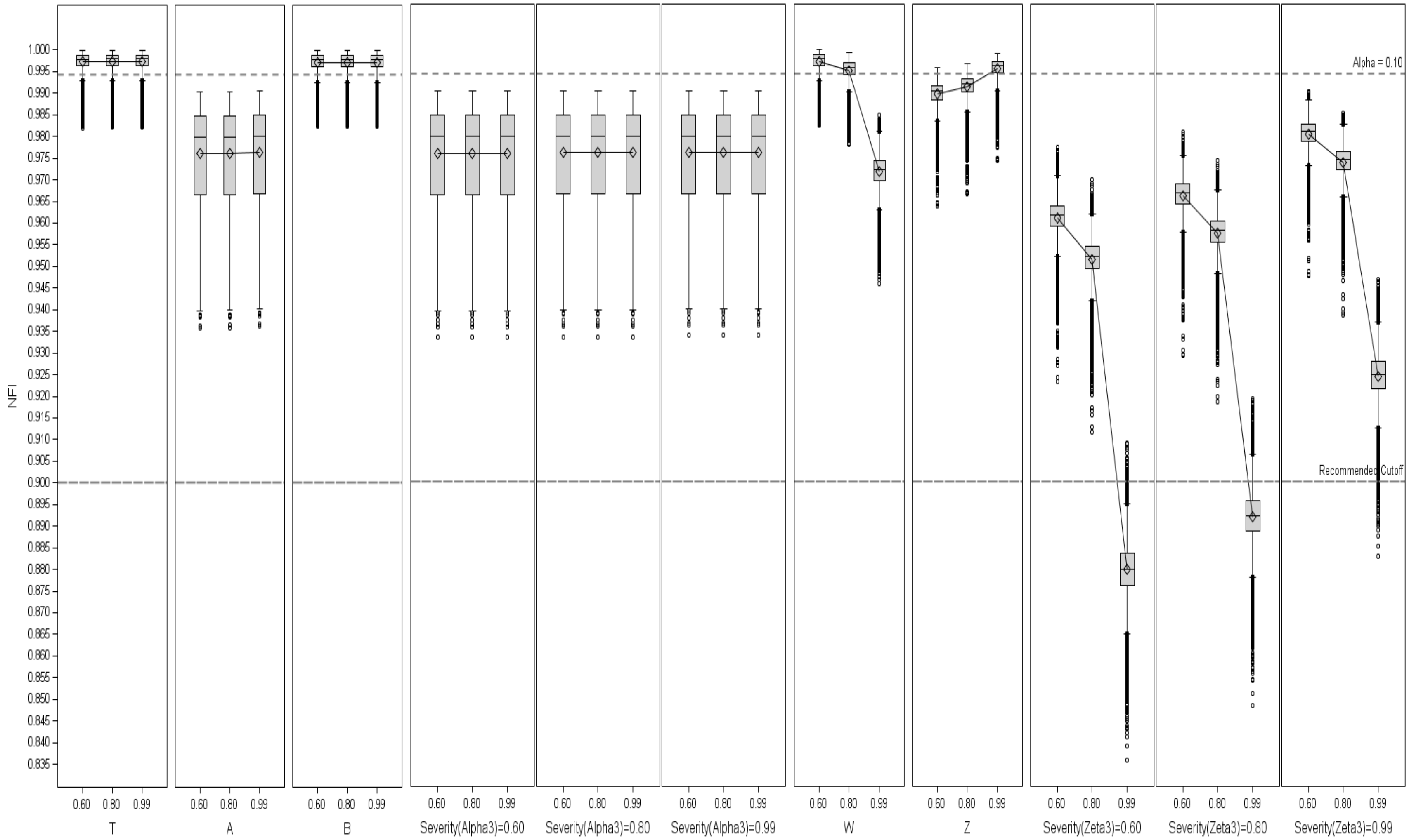


Figure 15: The effects of (misspecification×severity) on the performance of *NFI*.

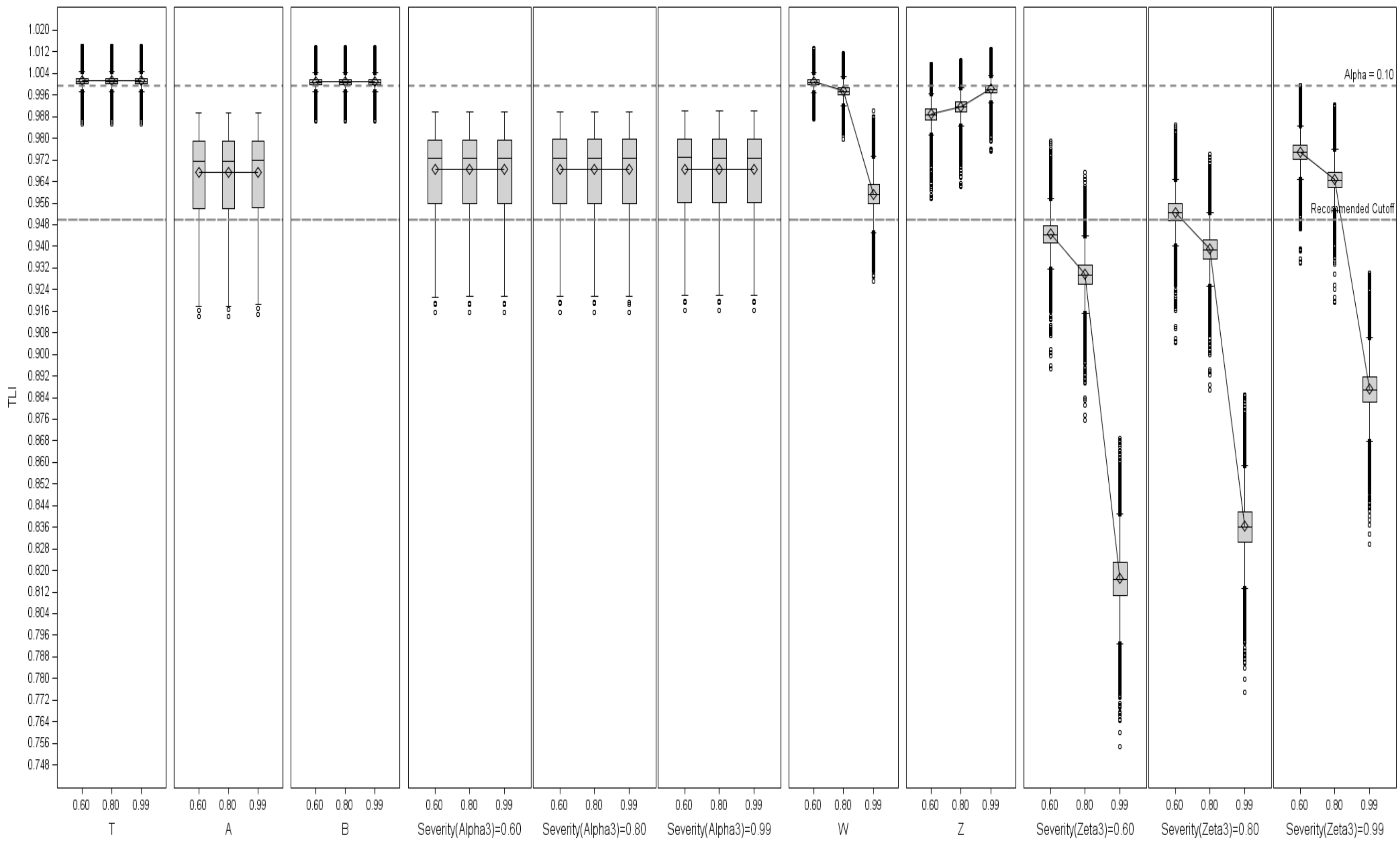


Figure 16: The effects of (misspecification×severity) on the performance of *TLI*.

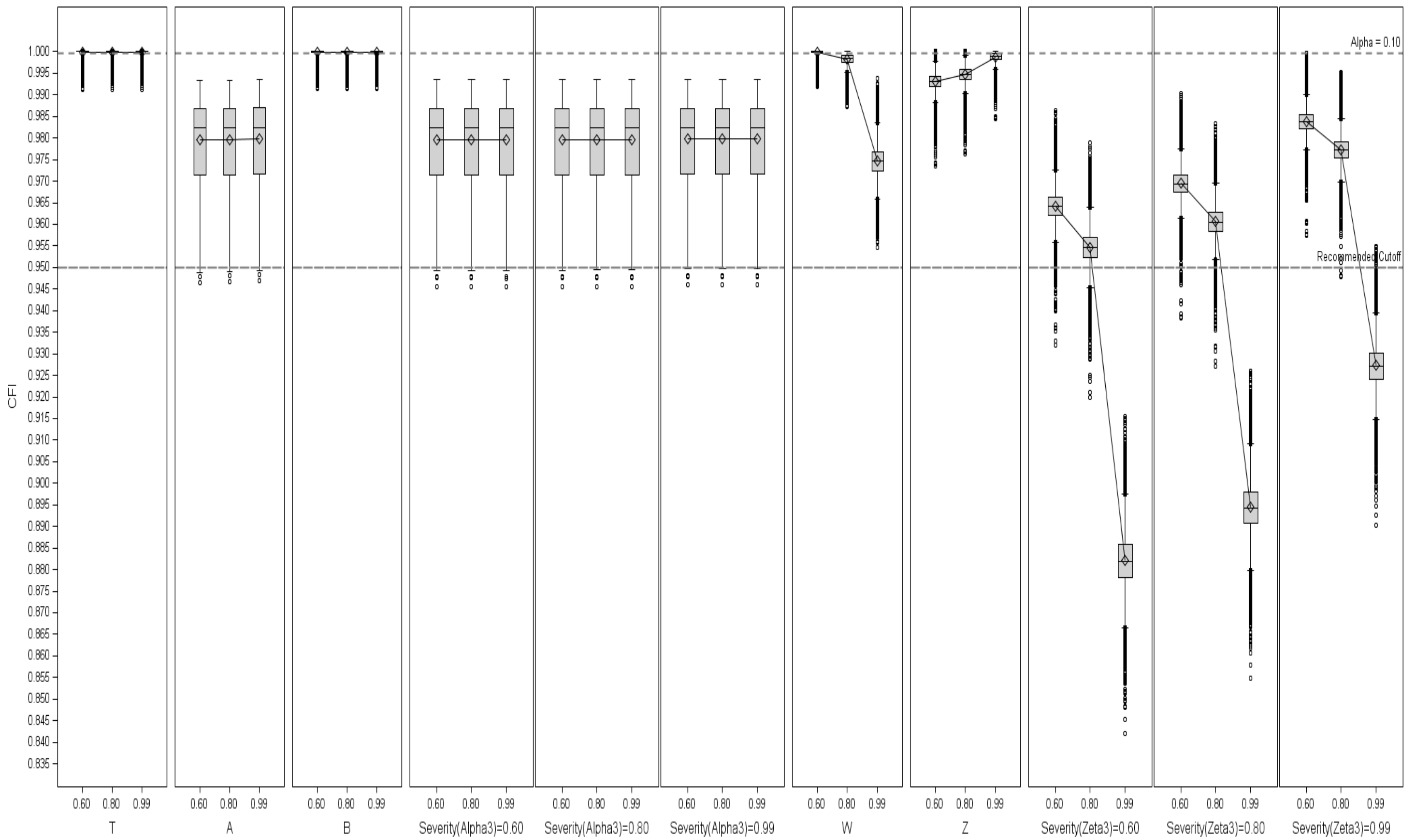


Figure 17: The effects of (misspecification×severity) on the performance of *CFI*.

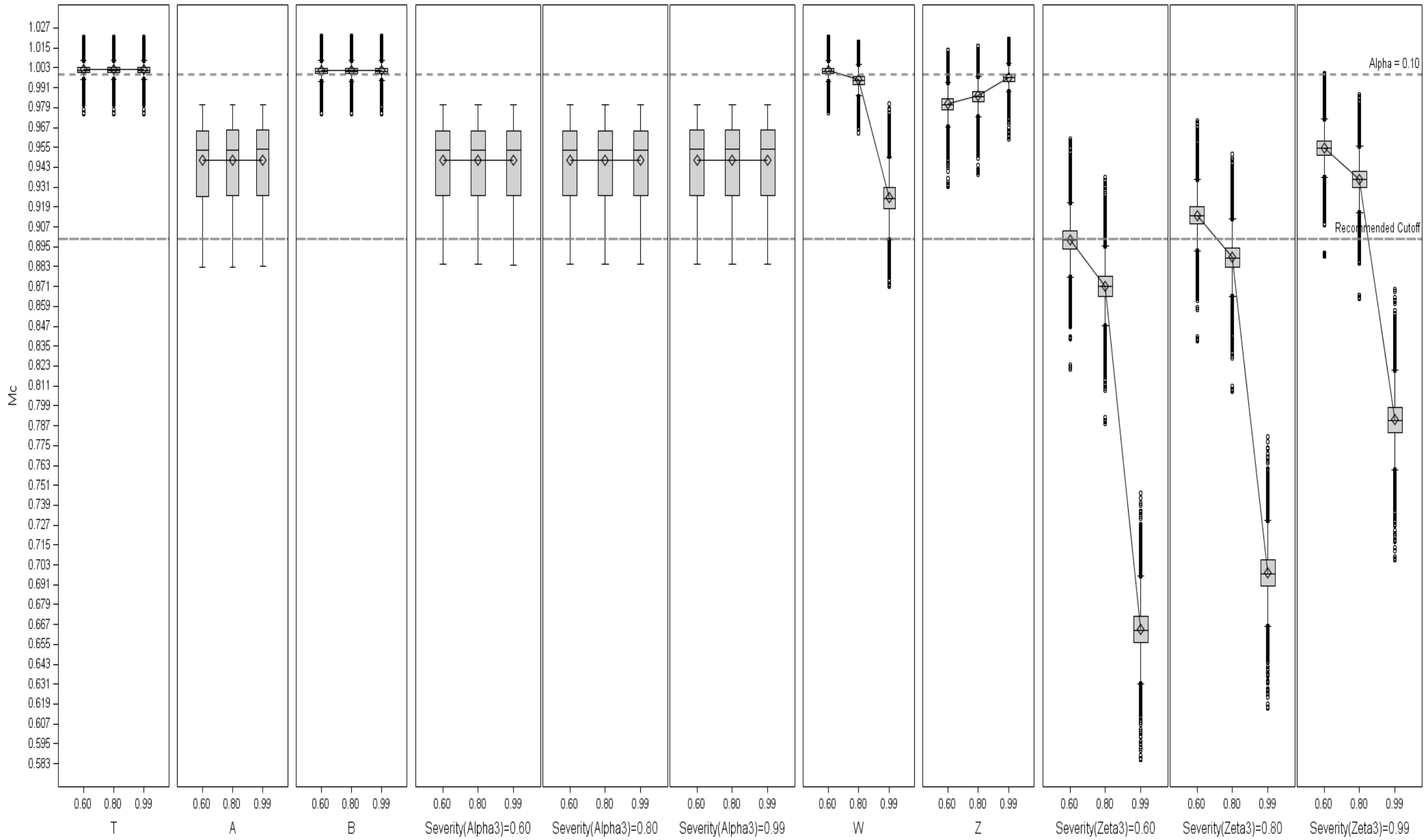


Figure 18: The effects of (misspecification×severity) on the performance of *Mc*.

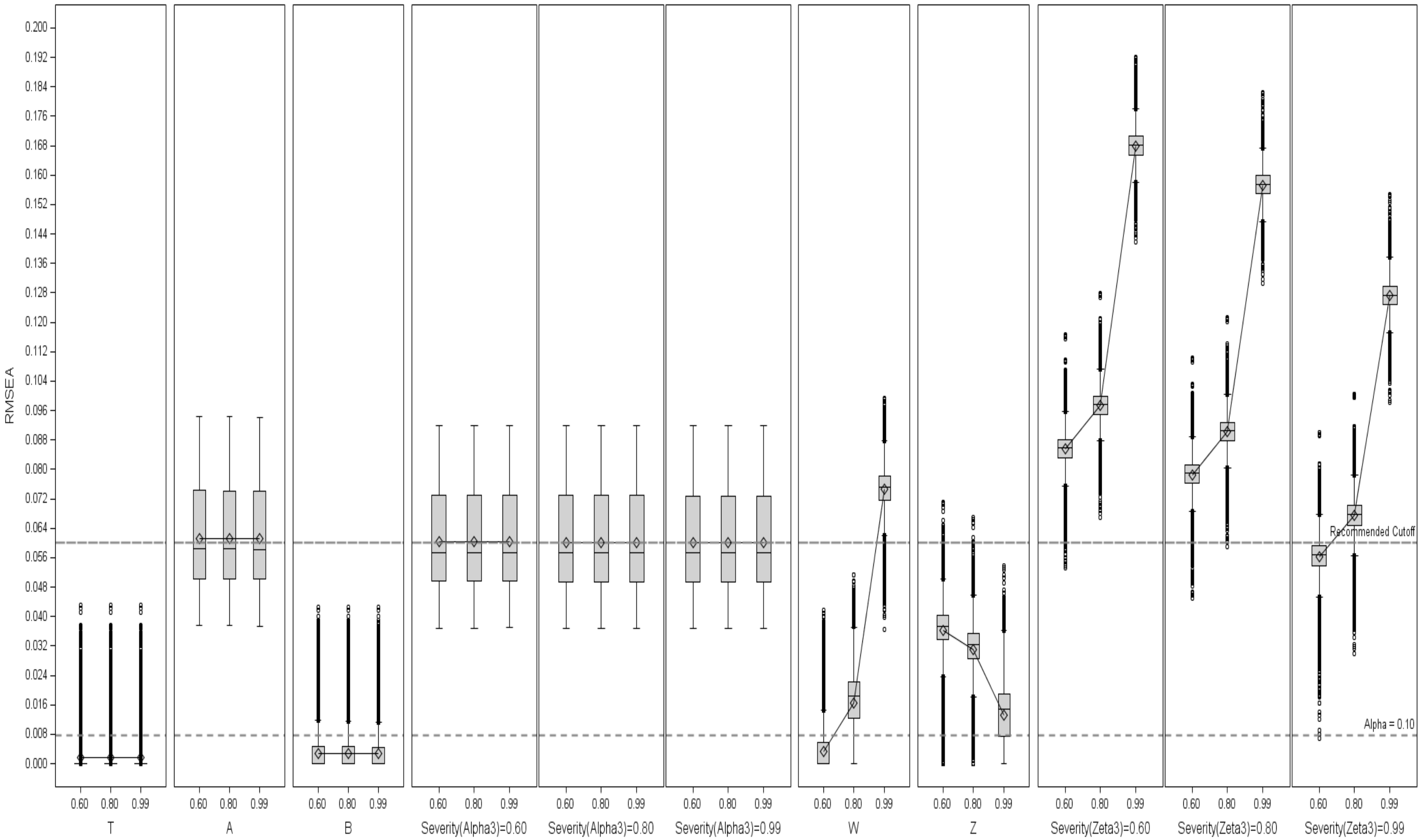


Figure 19: The effects of (misspecification×severity) on the performance of *RMSEA*.

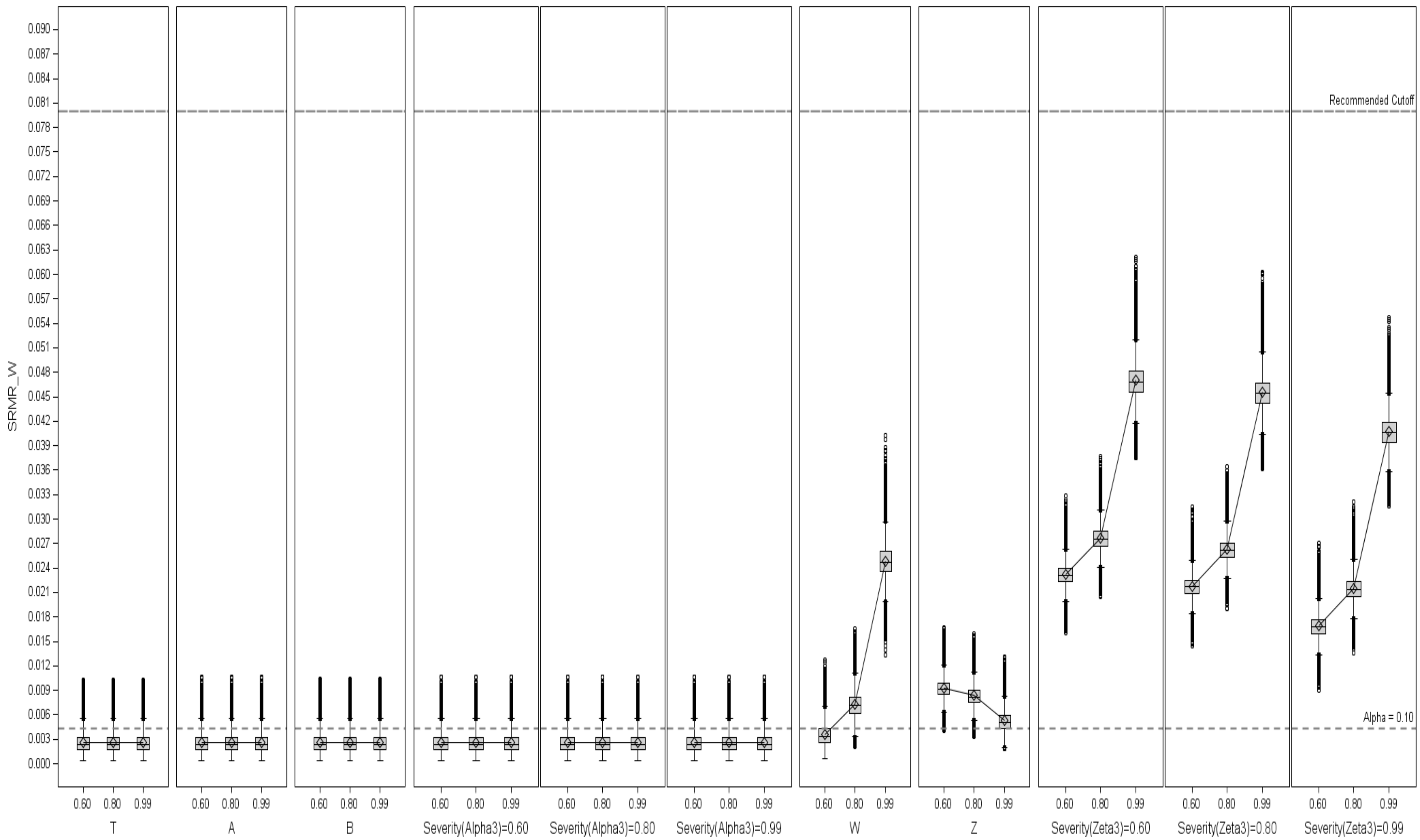


Figure 20: The effects of (misspecification×severity) on the performance of *SRMR_W*.

Effect of group size. As is shown in Figures 21 – 25, *NFI*, *TLI*, *CFI*, and *Mc* react similarly to group size when the misspecification involves either the marginal mean structure or both the marginal mean and the between-level covariance structures. The means of *NFI*, *TLI*, *CFI*, and *Mc* increase but the spread of those fit indices decrease with the increment of group size. The average and spread of *RMSEA*, however, both decrease when group size increases. When the misspecification occurs only in the between-level covariance structure, the spreads of all fit indices decrease with the increase of group size. The averages of those fit indices, however, change differently with the increase of group size. To be more specific, the average *NFI* increases whereas the averages of *TLI* and *Mc* decrease with the increase of group size. The averages of *CFI* and *RMSEA* do not change since these two fit indices are not significantly sensitive to group size.

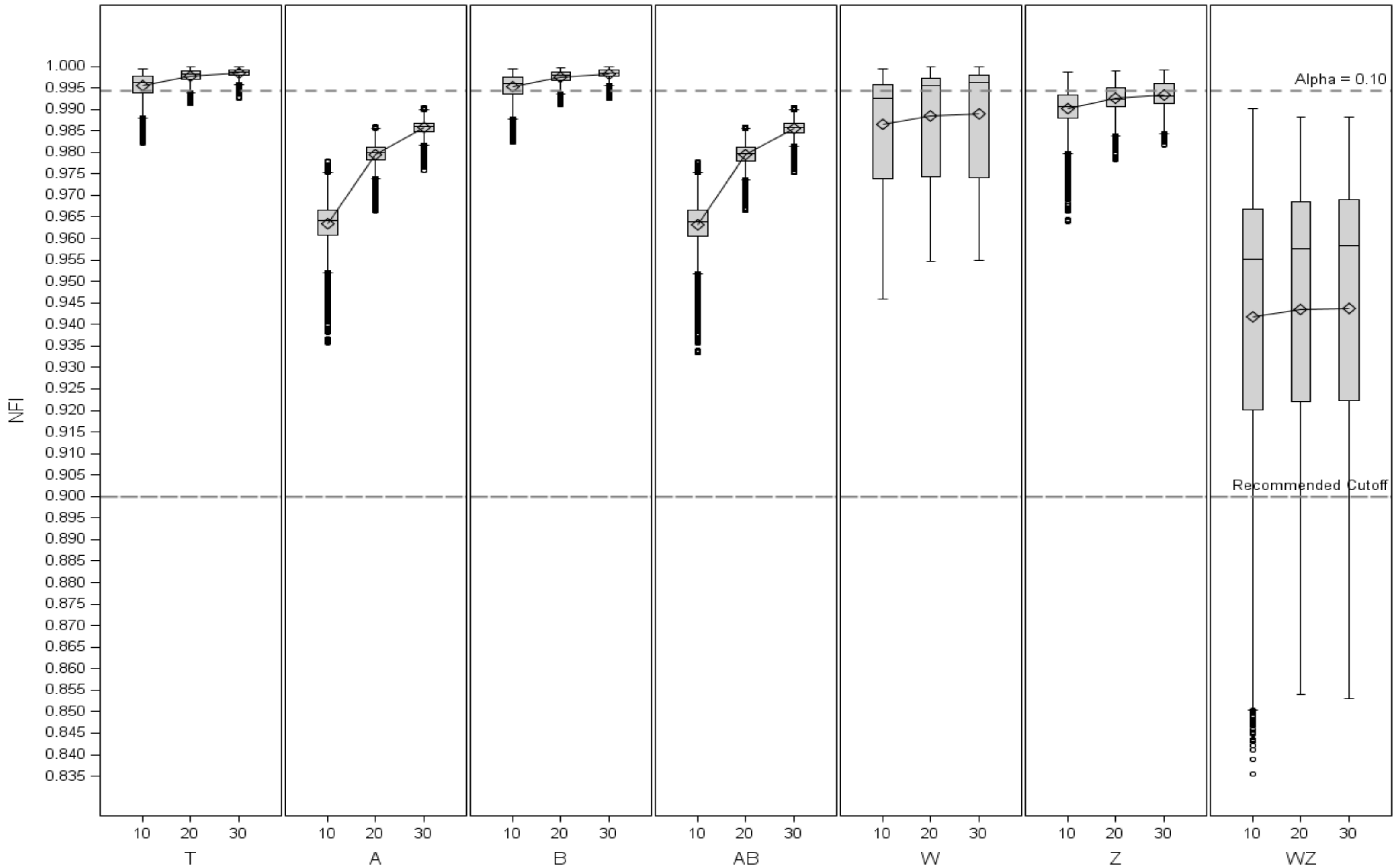


Figure 21: The effects of (misspecification×N_{gi}) on the performance of *NFI*.

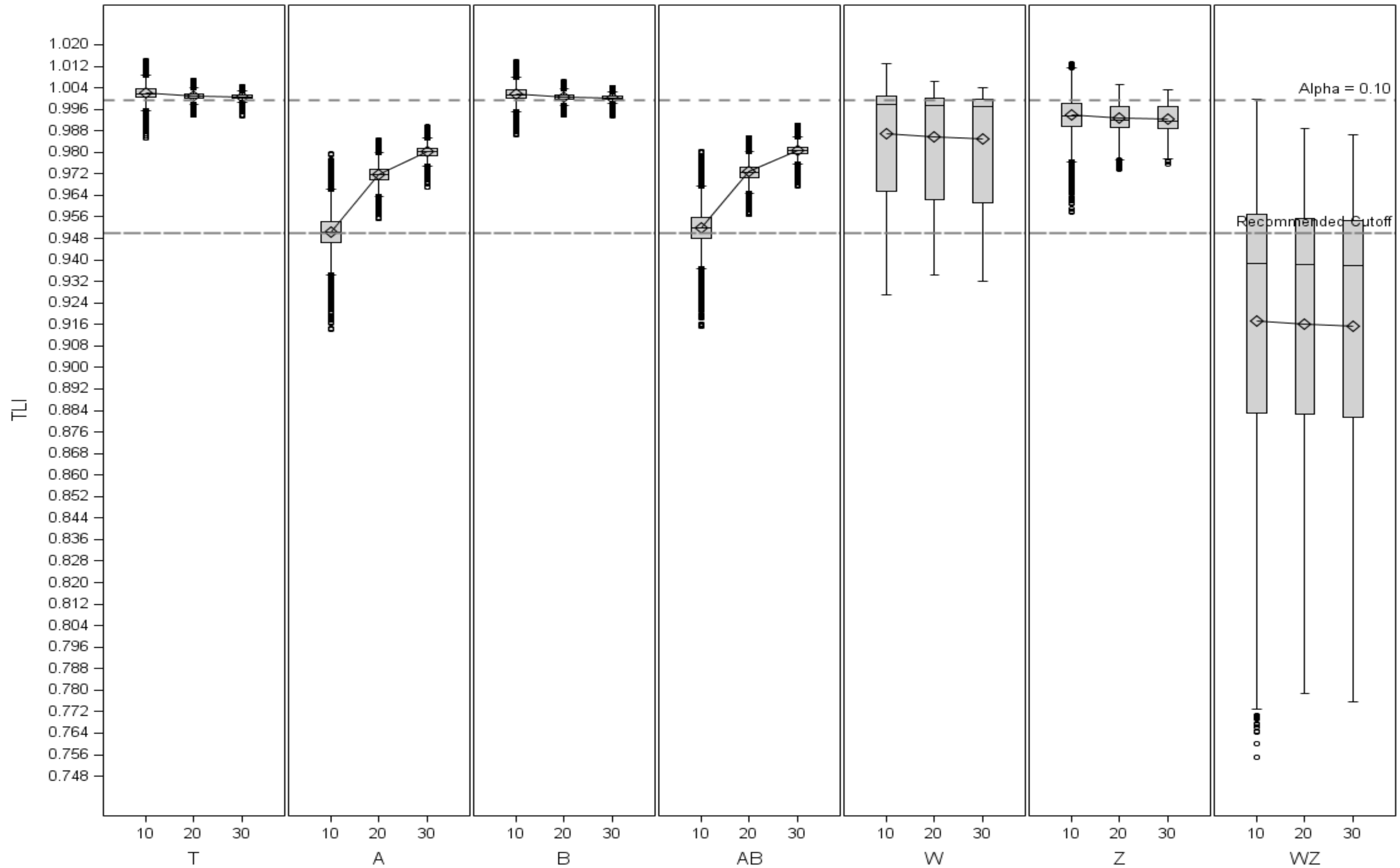


Figure 22: The effects of (misspecification×*Ngi*) on the performance of *TLI*.

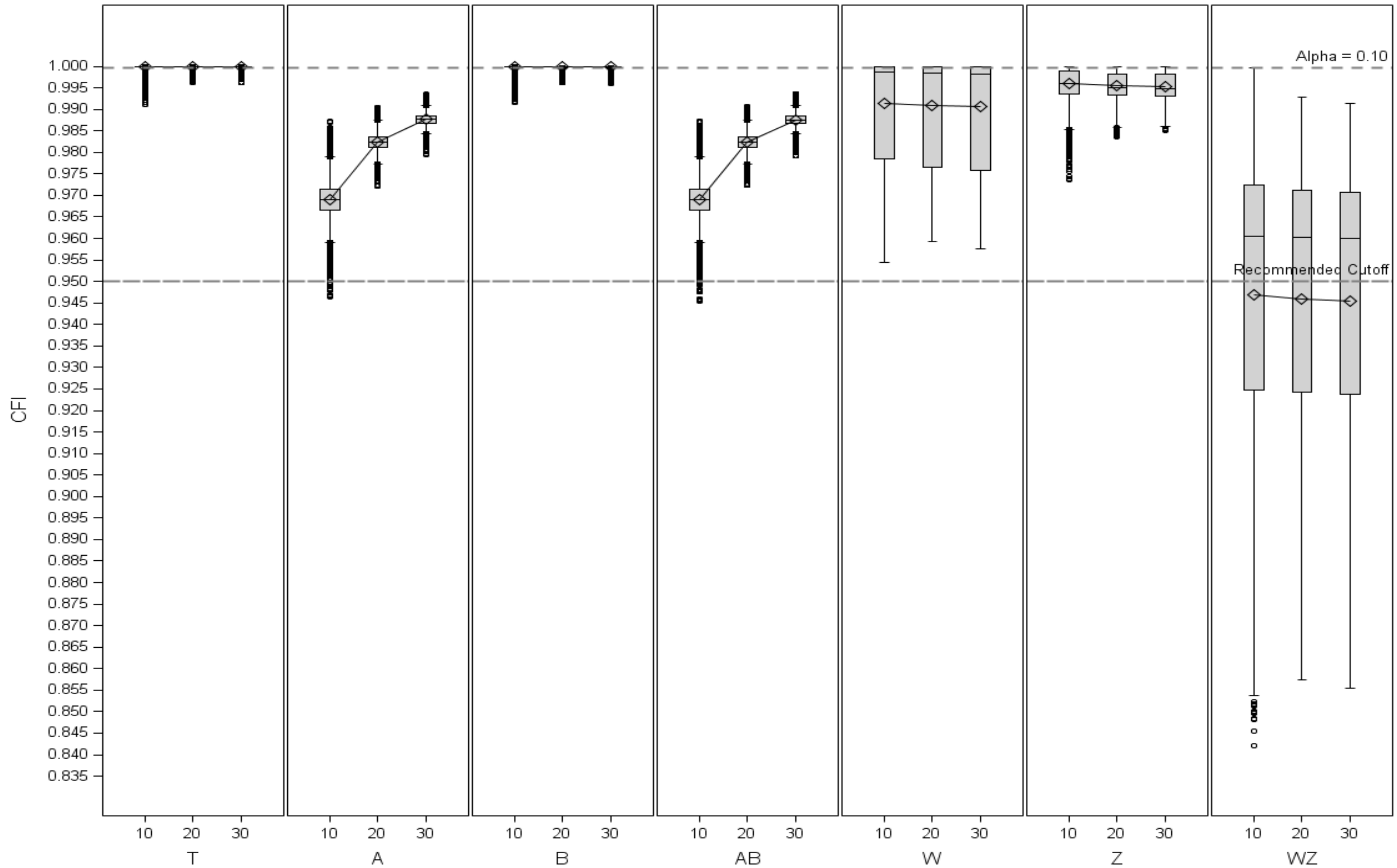


Figure 23: The effects of (misspecification×N_{gi}) on the performance of *CFI*.

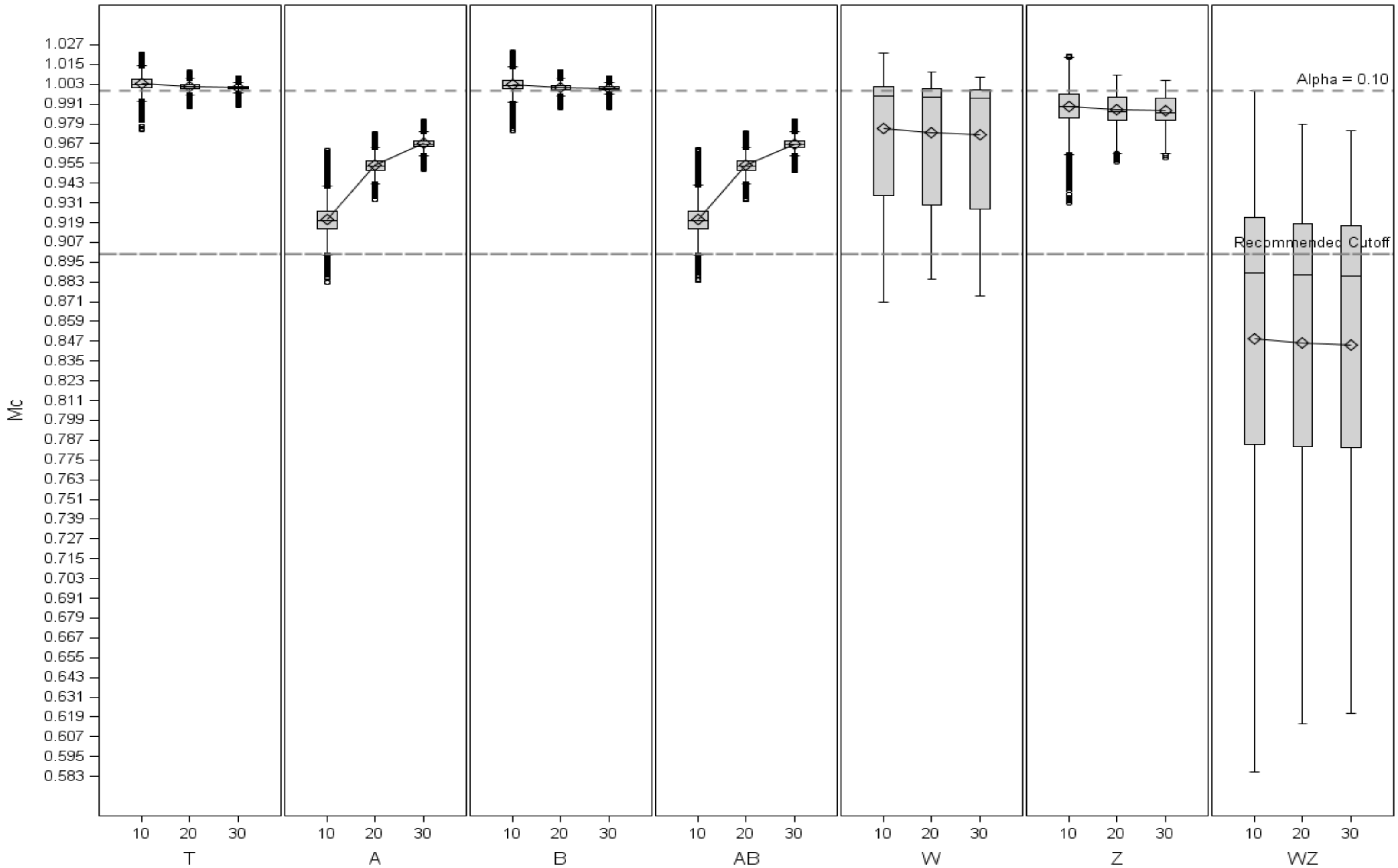


Figure 24: The effects of (misspecification×N_{gi}) on the performance of *Mc*.

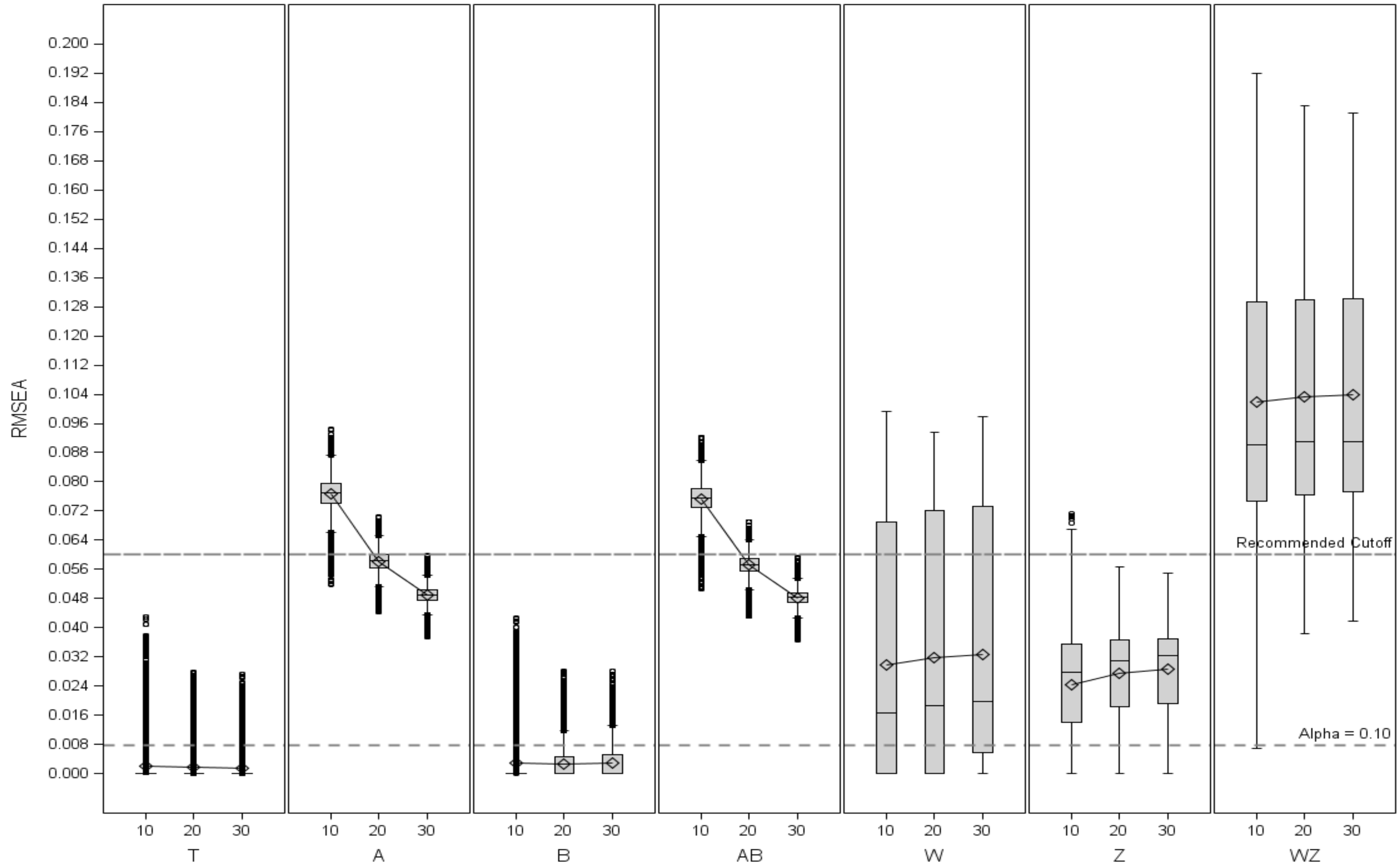


Figure 25: The effects of (misspecification×N_{gi}) on the performance of *RMSEA*.

Effect of number of groups. As is shown in Figures 26 – 28, *TLI* and *Mc* respond to the group-level sample size in the same way. When the number of groups increases, both the means and the standard deviations of *TLI* and *Mc* decrease if the model is misspecified in the between-level covariance structure, implying that they have more power to reject the misspecified model with the increase of group number. The standard deviation of *NFI* also decreases with the increment of the group-level sample size when the model is misspecified in the between-level covariance structure and in the within-level residual structure. However, its mean increases with the growth of group numbers in the data, suggesting that it is more likely to commit Type II errors when more independent groups are included in the data.

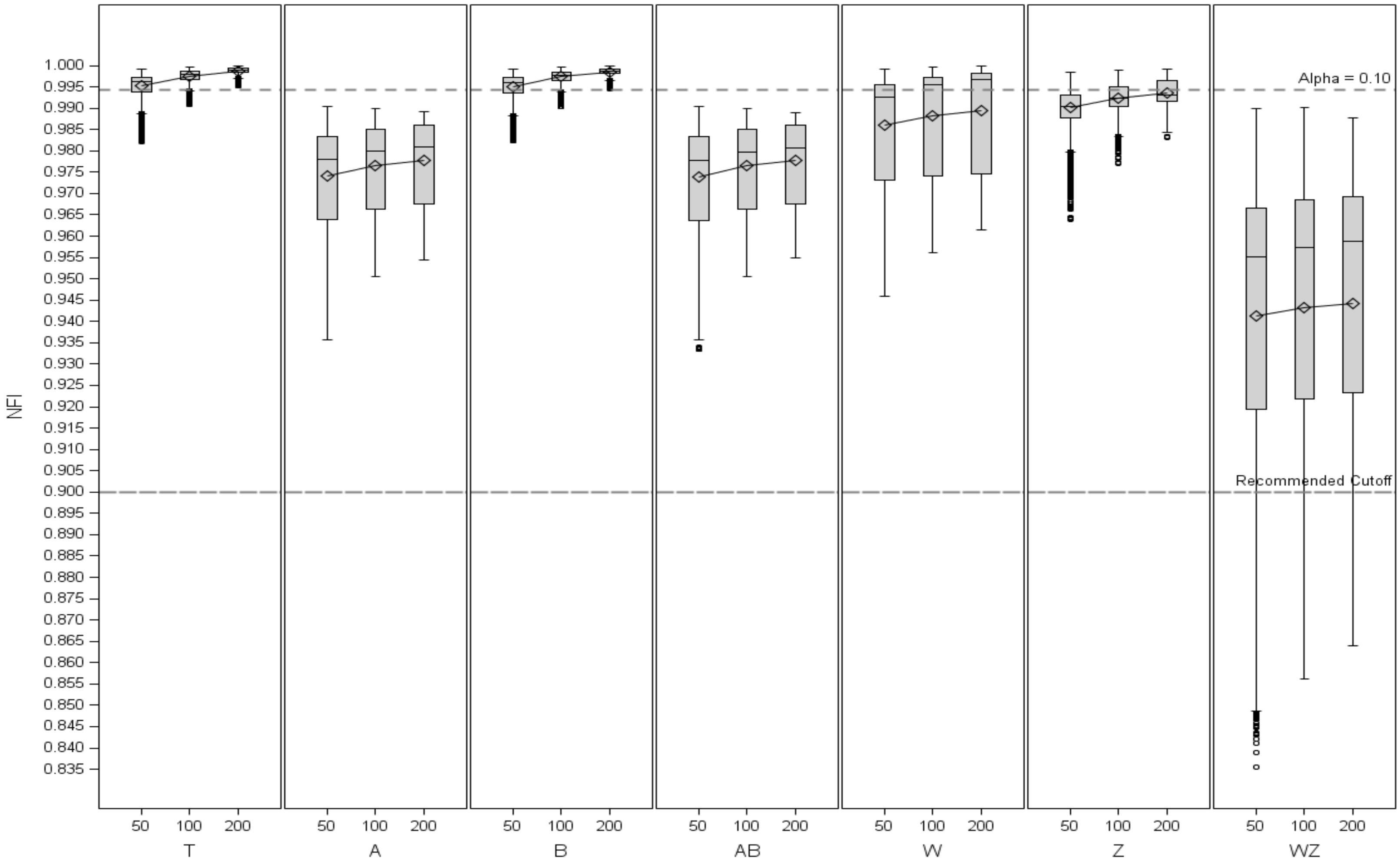


Figure 26: The effects of (misspecification×N_g) on the performance of *NFI*.

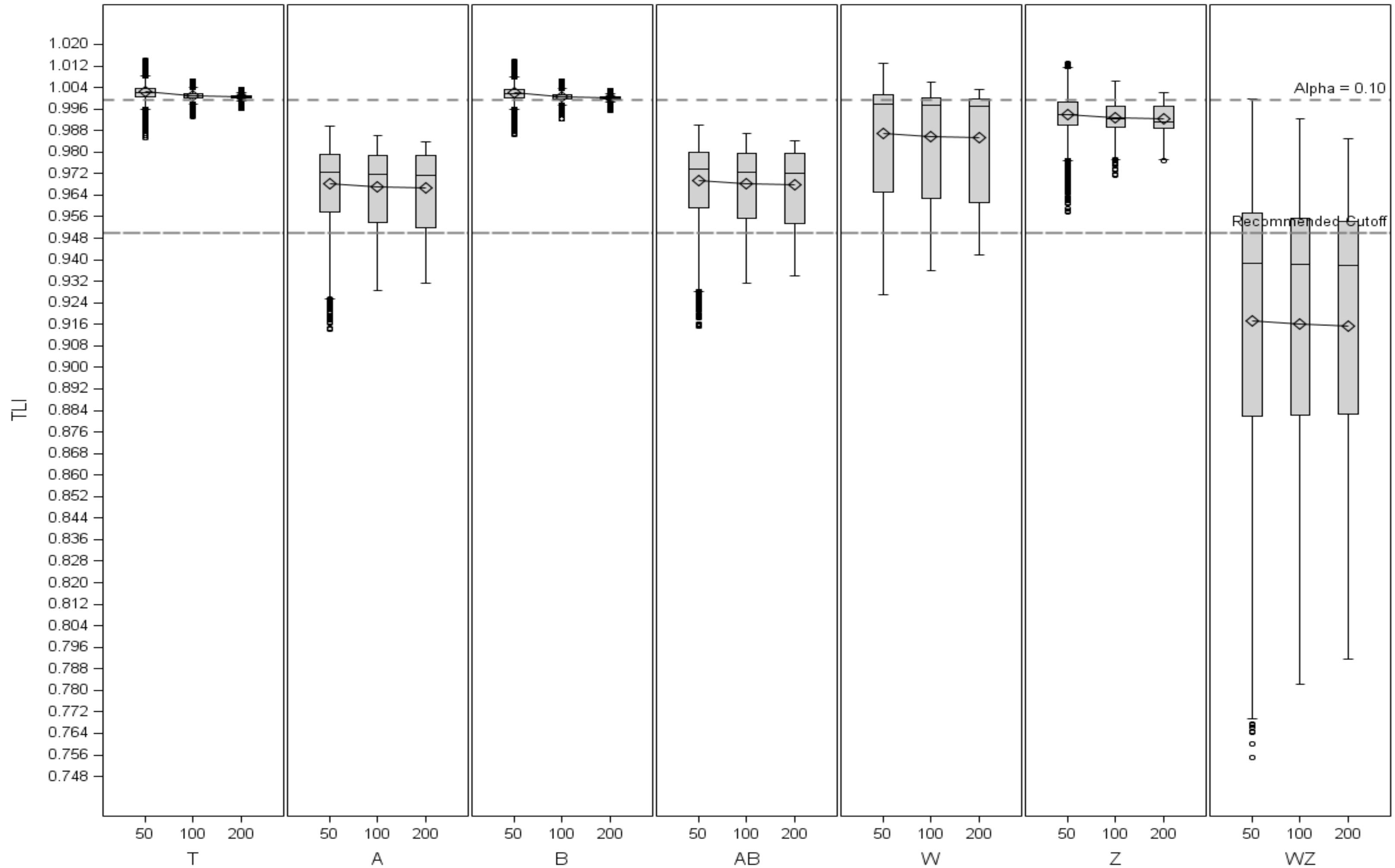


Figure 27: The effects of (misspecification×N_g) on the performance of *TLI*.

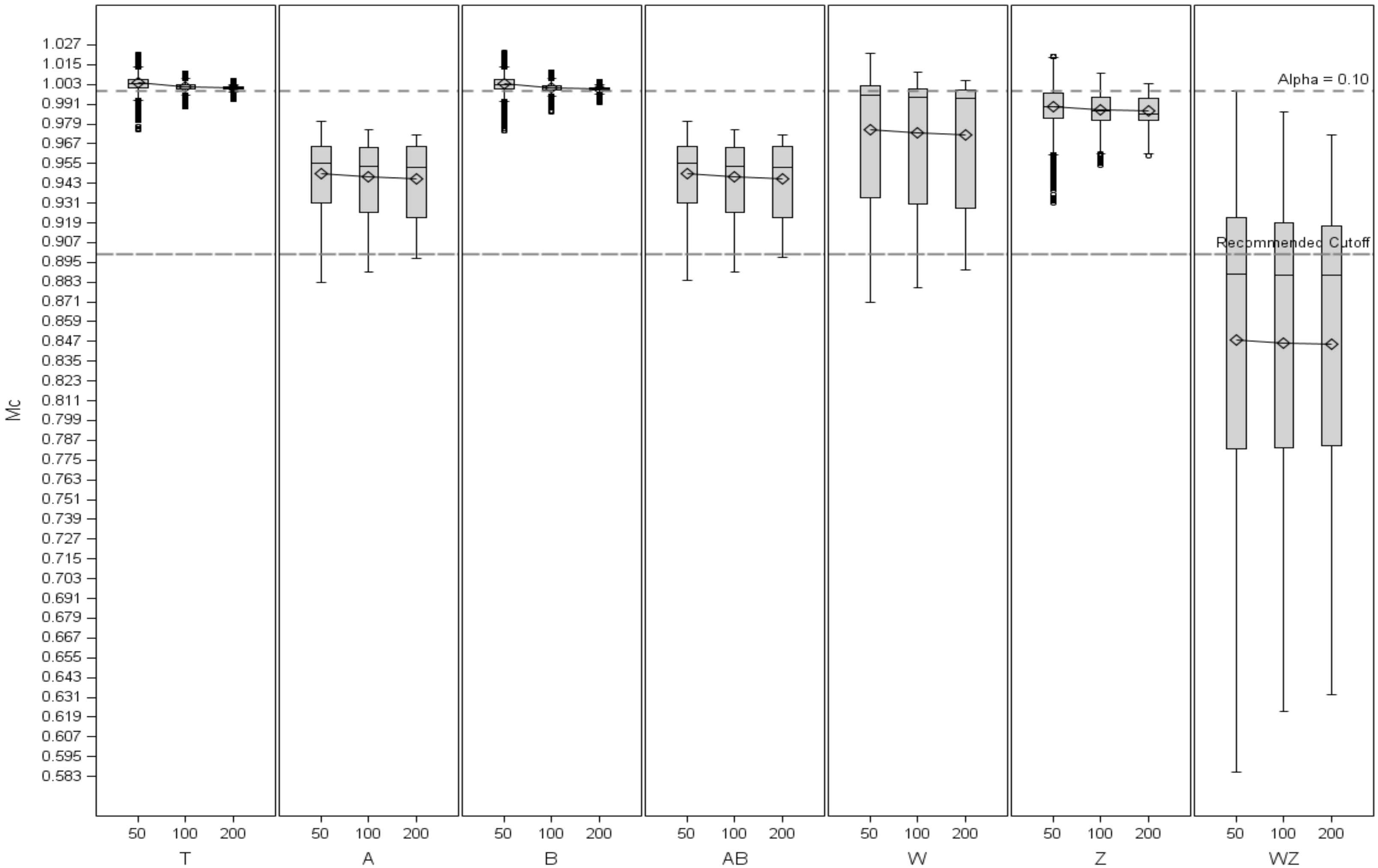


Figure 28: The effects of (misspecification×N_g) on the performance of *Mc*.

Effect of intraclass correlation coefficient. *SRMR* is the only index that is sensitive to the intraclass correlation coefficient when the between-level structure (i.e., the marginal mean and the covariance structures) of the linear-linear piecewise multilevel latent growth curve model is misspecified. When the intraclass correlation coefficient increases, both the means and standard deviations of *SRMR* declines (see Figure 29), implying that it is more likely to commit Type II errors with higher intraclass correlation coefficient.

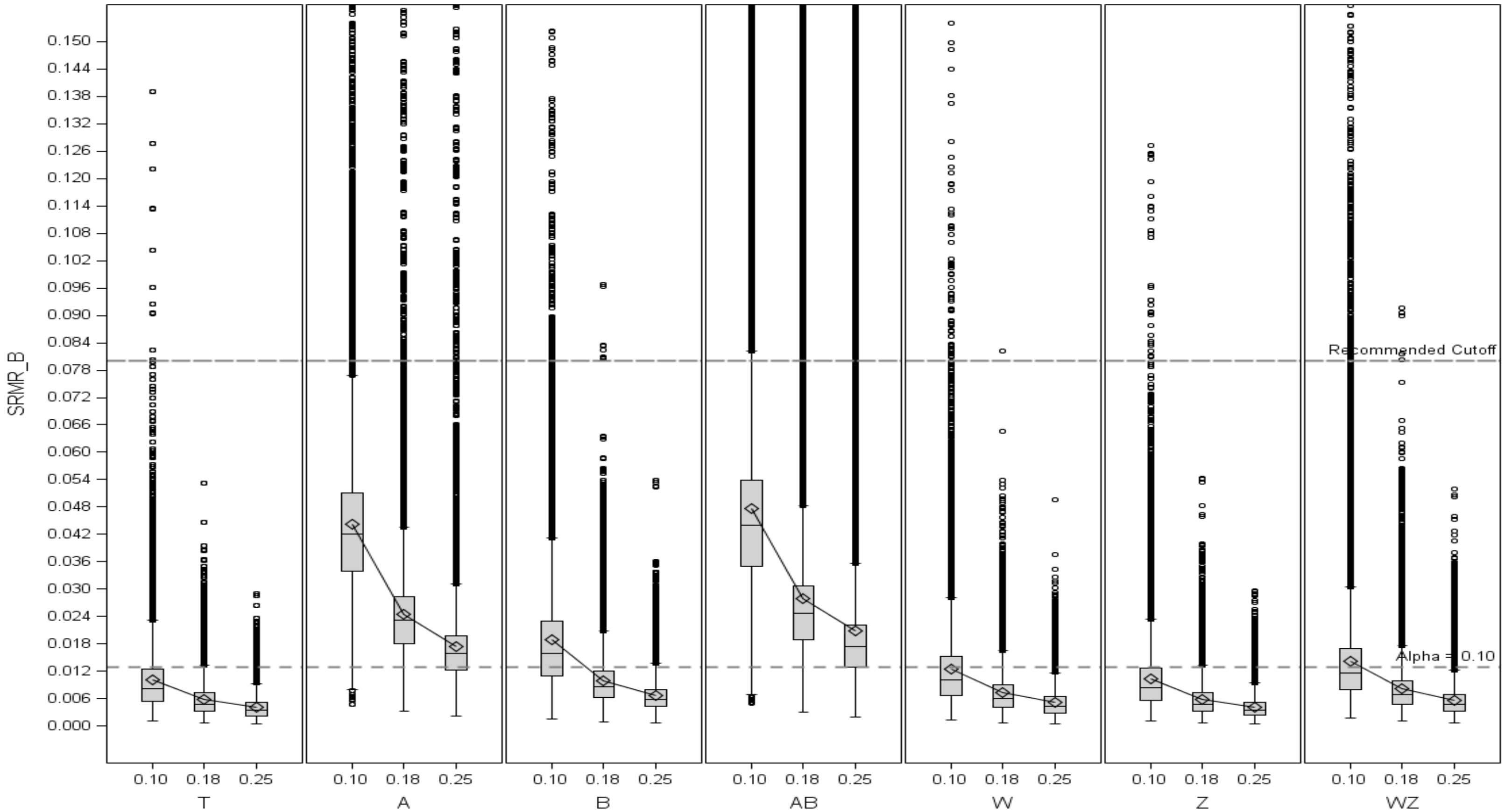


Figure 29: The effects of (misspecification×ICC) on the performance of *SRMR_B*[32].

[32] The values of *SRMR_B* were truncated to better demonstrate the relationship.

4.1.4 Correlation analyses[33]

Several rules of thumb have been proposed for the interpretation of the Pearson correlation coefficient. This study adopts the one proposed by Hinkle, Wiersma, and Jurs (2003), which divides the correlation coefficient into five categories: values ranging from 0.90 to 1.0 are regarded as very high correlation, values varying from 0.70 to 0.90 are assumed to be high, values oscillating from 0.50 to 0.70 are considered as moderate, values bounded within 0.30 and 0.50 represent low correlations, and values smaller than 0.30 are considered little and negligible.

Two fit indices are expected to have high correlations if they react to the same type of misspecification in the same or similar way. On the contrary, they are likely to have low to moderate correlations if they respond differently to a type of misspecification. In addition, since larger values of *NFI*, *TLI*, *CFI*, and *Mc* represent that the model has better chances to be accepted whereas smaller values of *RMSEA* and *SRMR* means better model fit, *NFI*, *TLI*, *CFI*, and *Mc* are expected to be negatively correlated with *RMSEA* and *SRMR*. Moreover, *NFI*, *TLI*, *CFI*, and *Mc* are expected to be positively correlated with each other, as is *RMSEA* and *SRMR*.

When the model is misspecified in the marginal mean structure (see Figure 30), *NFI*, *TLI*, *CFI*, and *Mc* follow similar distributions, all of which have three modes and are negatively skewed. *RMSEA*, however, follows a distribution that is similar to those of *NFI*, *TLI*, *CFI*, and *Mc* but in a converse direction. Therefore, *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are highly correlated with each other, with *RMSEA* being negatively correlated with *NFI*, *TLI*, *CFI*, and *Mc*. In contrast to the multi-modal

[33] Correlation analysis might not be proper for cases where the bivariate relations are not linear.

distributions of *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA*, *SRMR* follows a uni-modal distribution which is positively skewed, making it being barely correlated with all other fit indices. In other words, *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* react similarly to the

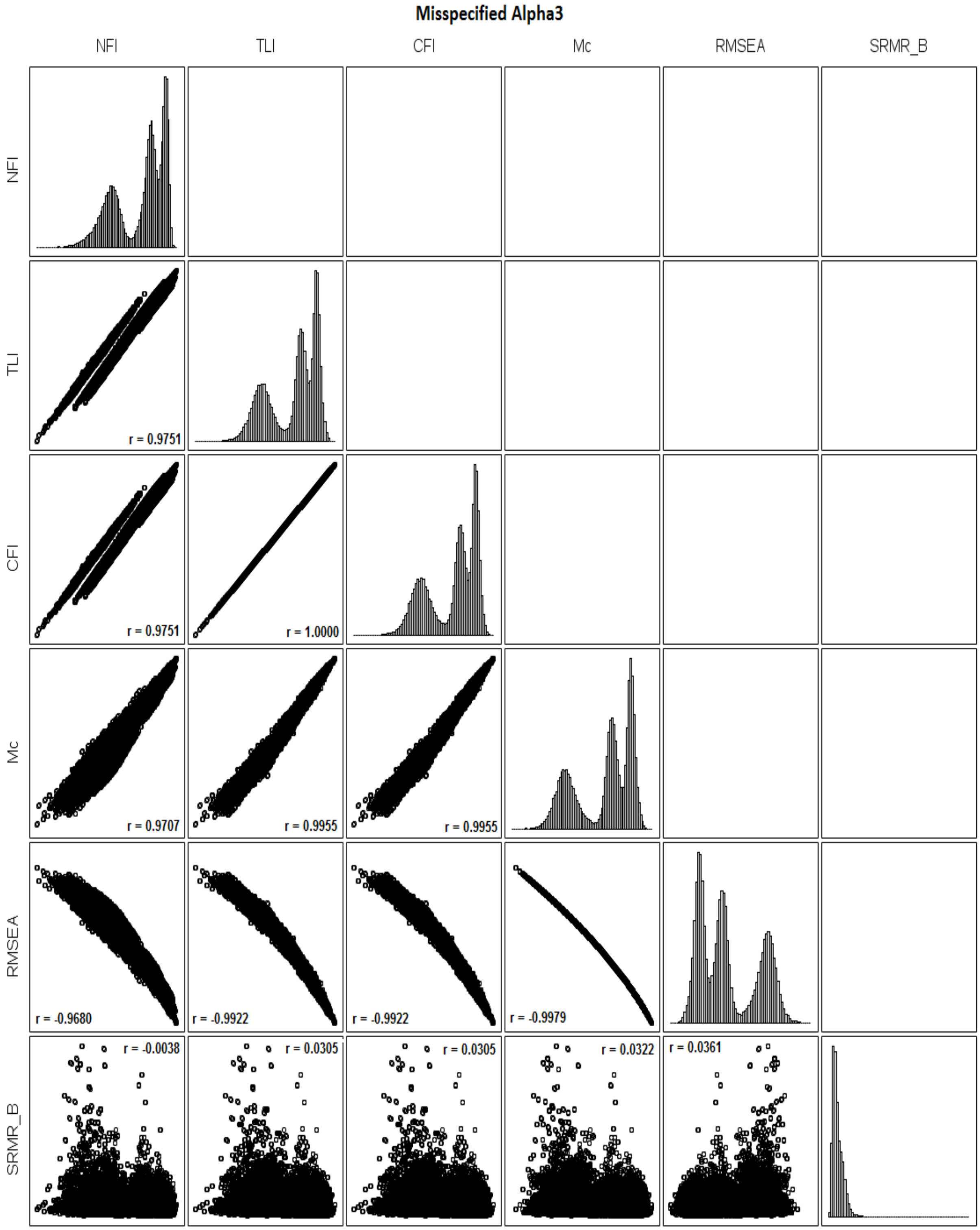


Figure 30: Correlations of the fit indices for models misspecified in the marginal mean structure.

misspecification in the marginal mean structure of the linear-linear piecewise multilevel latent growth curve model. Nevertheless, *SRMR* behaves differentially from other fit indices in detecting misspecifications in the marginal mean structure.

When the model is misspecified in the between-level covariance structure (see Figure 31), *TLI* and *Mc* are positively highly correlated with each other and *CFI* and *RMSEA* are negatively highly correlated with each other. The former pair of fit indices follows a uni-modal bell-shaped distribution that is slightly positively skewed. The latter pair of fit indices, however, both follows distributions that are uni-modal and highly skewed. The distributions of *NFI* and *SRMR* look different from those of all other fit indices, making them minimally to moderately correlated with other fit indices. In other words, *TLI* and *Mc*, and *CFI* and *RMSEA* may work interchangeably in detecting the misspecification in the between-level covariance structure of linear-linear piecewise multilevel latent growth curve models. *SRMR*, however, works differentially from all other practical fit indices and cannot be a substitute for other fit indices.

When the model is misspecified in both the marginal mean structure and the between-level covariance structure (see Figure 32), all fit indices follow distributions that are identical to those when the misspecification involves only the marginal mean structure. In other words, *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are exchangeable in detecting misspecifications in both the marginal mean and the between-level covariance structures. However, *SRMR* responds differently to these misspecifications when compared to other fit indices.

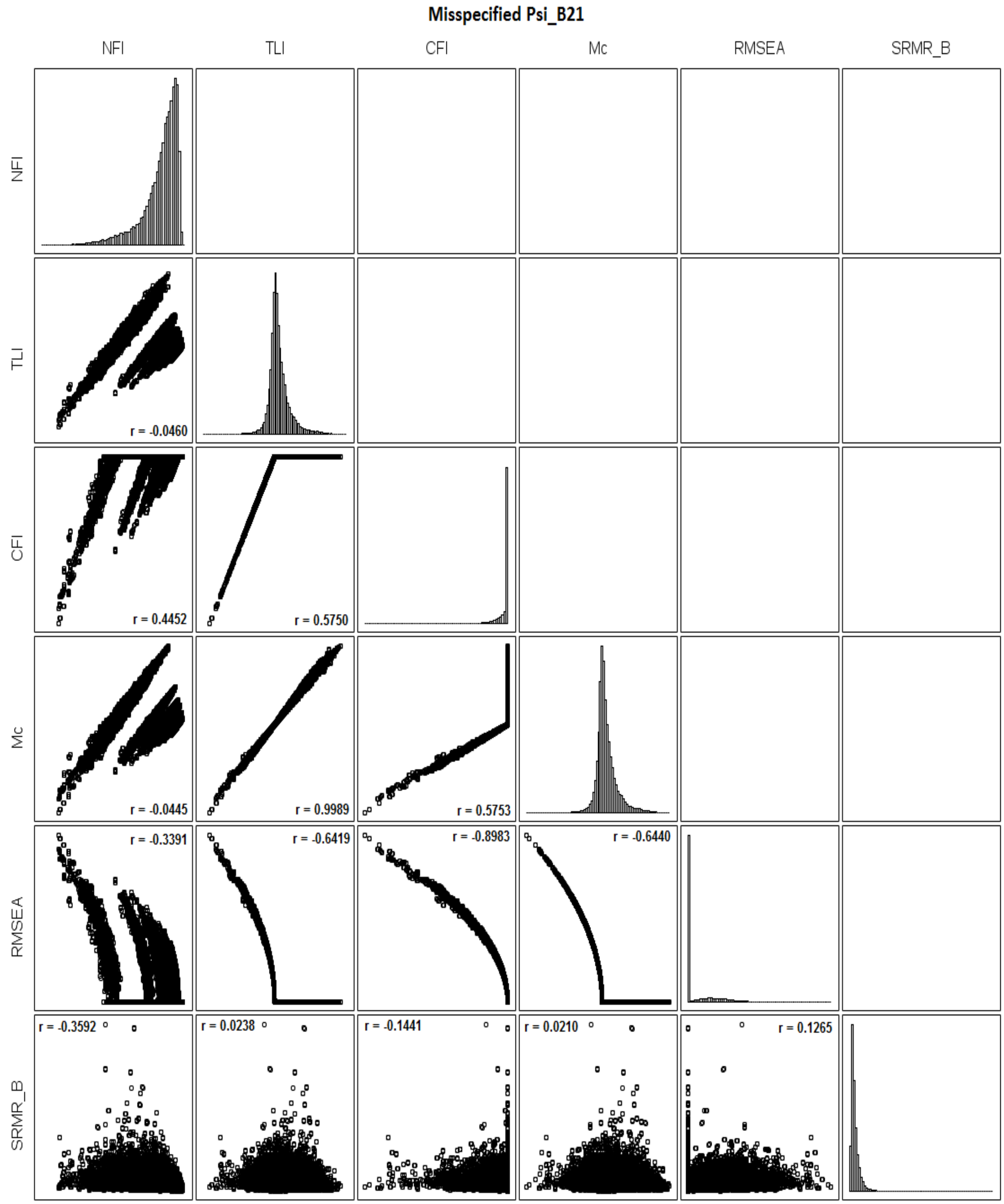


Figure 31: Correlations of the fit indices for models misspecified in the between-level covariance structure.

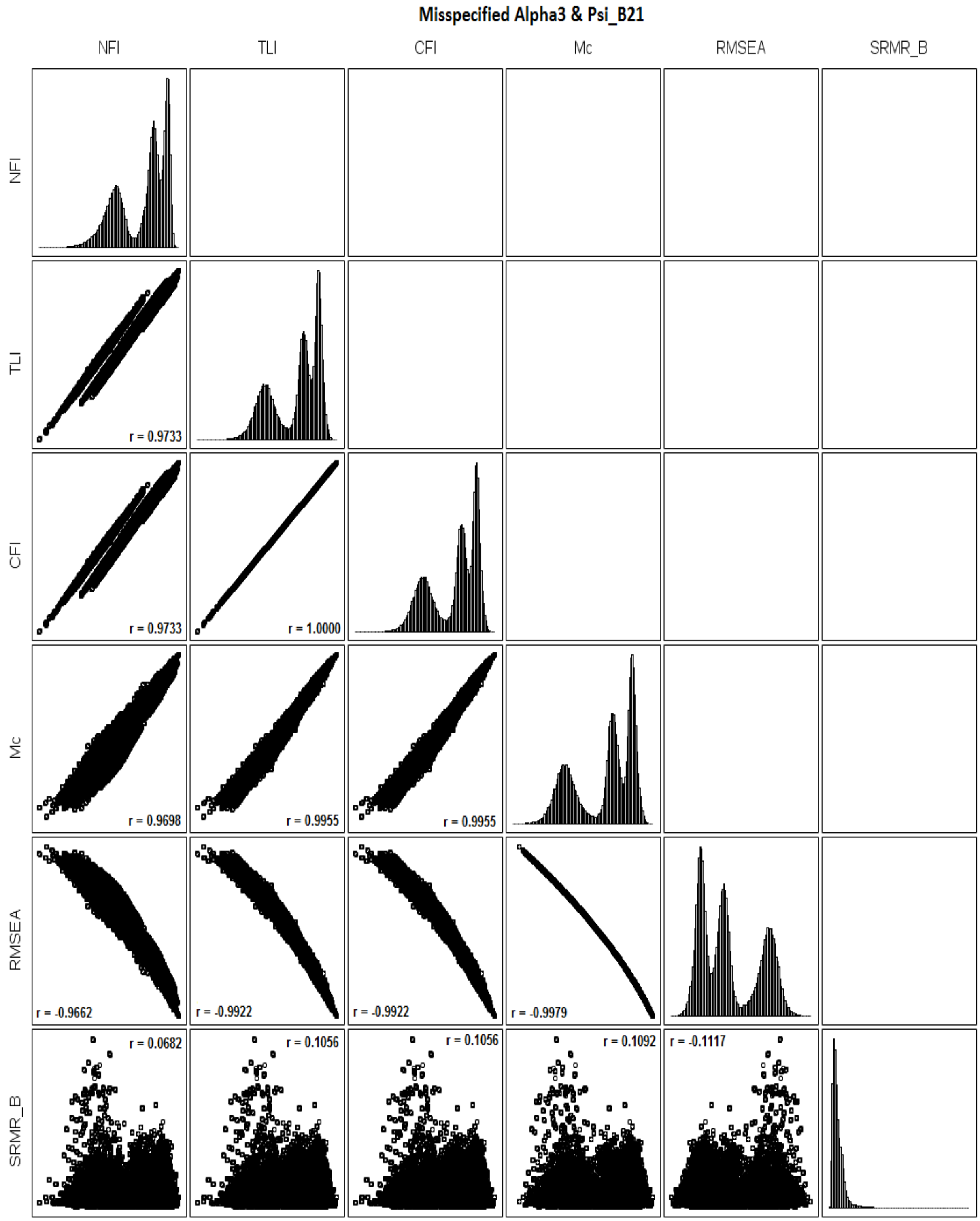


Figure 32: Correlations of the fit indices for models misspecified in both the marginal mean and the between-level covariance structures.

When the model is misspecified in the within-level covariance structure (see Figure 33), all fit indices are highly correlated with each other. When looking at their respective distributions, it is found that *NFI*, *TLI*, *CFI*, and *Mc* all have bi-modal distributions, with the major mode appearing on the right. In addition, the distributions of *TLI* and *Mc* are almost identical to each other, implying that these two fit indices may be interchangeable in detecting the misspecification in the within-level covariance structure. In contrast to *NFI*, *TLI*, *CFI*, and *Mc*, *RMSEA* and *SRMR* follow multi-modal distributions whose major mode appears on the left, suggesting that they work differentially from other fit indices in detecting this type of model misspecification.

When the model is misspecified in the within-level residual structure (see Figure 34), all fit indices are moderately to highly correlated with each other. Among all fit indices, *TLI* and *Mc* follow almost identical distributions, making them interchangeable in detecting this type of misspecification. All other fit indices, although highly correlated with each other, followed different distributions and may not respond in the same way to the misspecified within-level residual structure.

When the model is misspecified in both the within-level covariance and residual structures (see Figure 35), *NFI*, *TLI*, *CFI*, and *Mc* follow nearly identical distributions and are highly correlated with each other. In other words, these fit indices may be interchangeable in detecting misspecifications on the within-level structure of linear-linear piecewise multilevel latent growth curve models. Although *SRMR* is highly correlated with all other fit indices, it follows a different distribution, suggesting that it might work differently from other fit indices in evaluating models misspecified in the within-level structure.

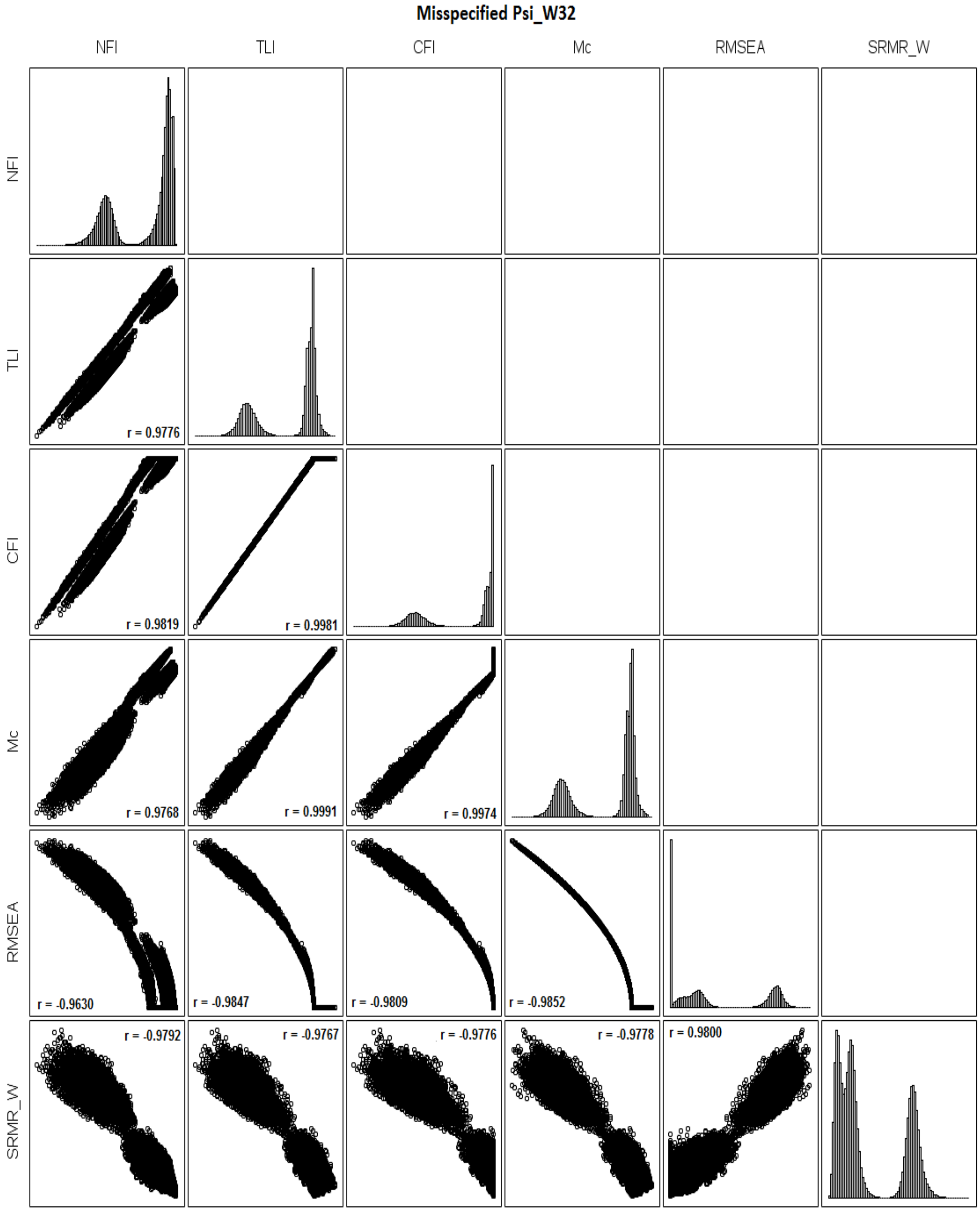


Figure 33: Correlations of the fit indices for models misspecified in the within-level covariance structure.

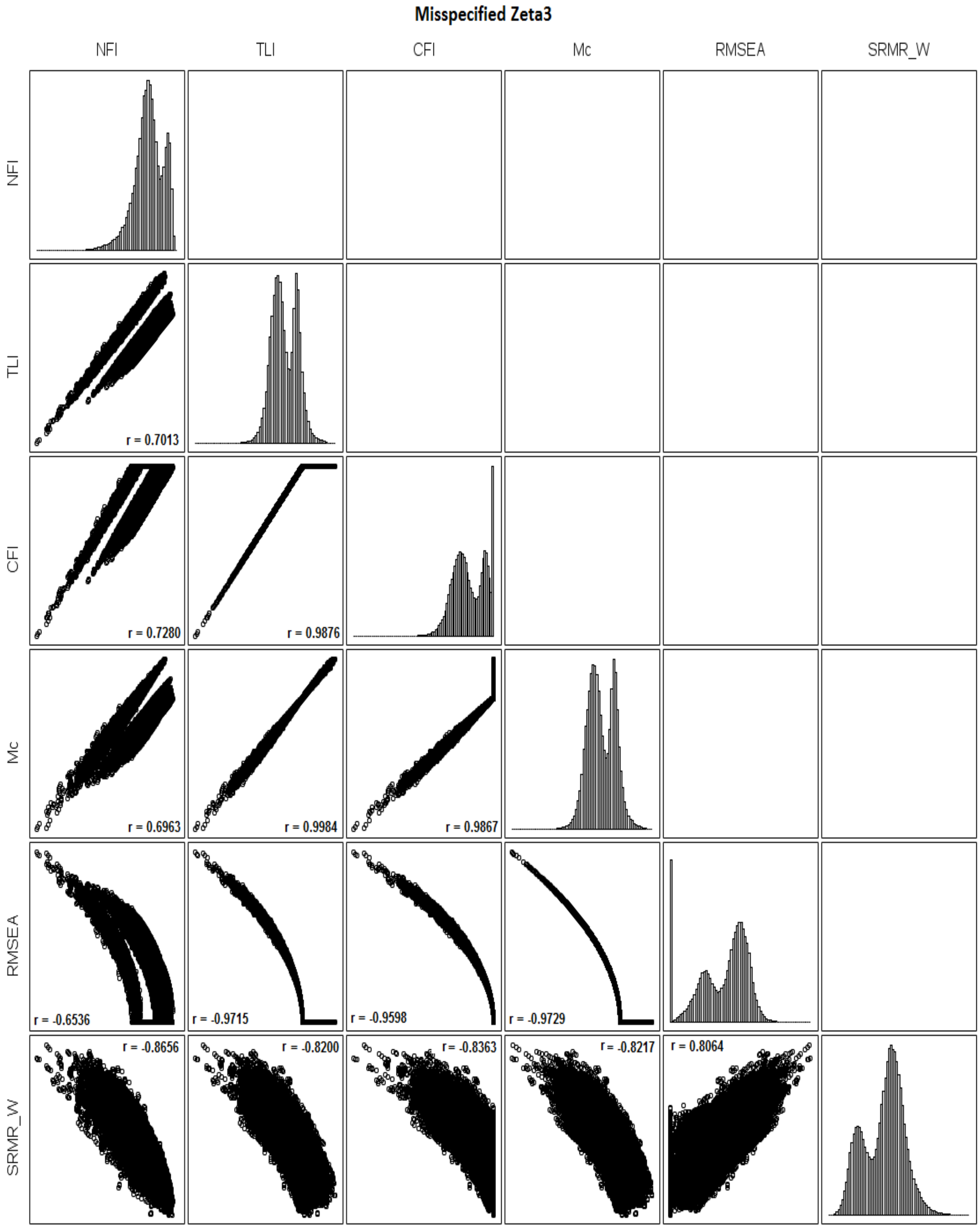


Figure 34: Correlations of the fit indices for model misspecified in the within-level residual structure.

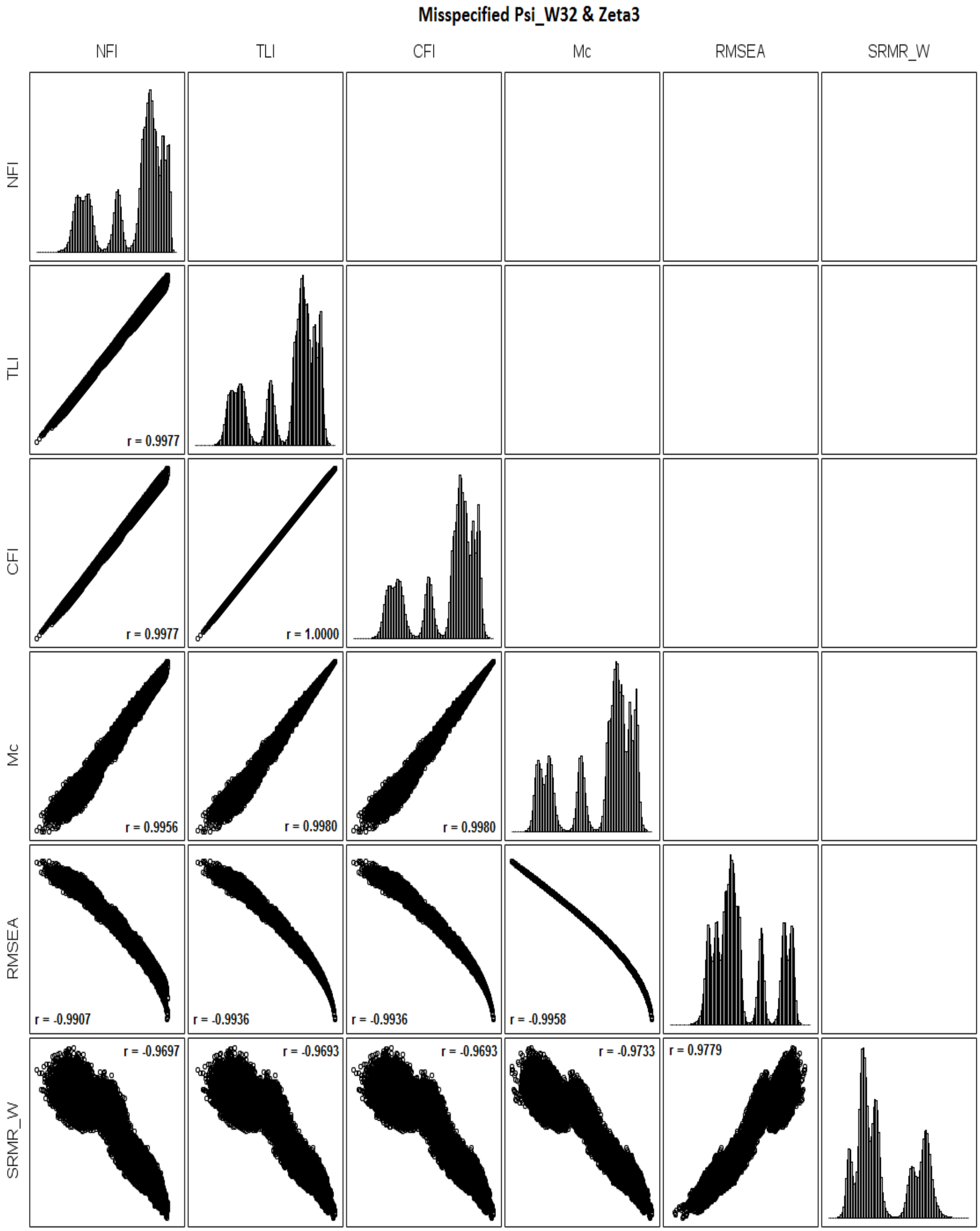


Figure 35: Correlations of the fit indices for models misspecified in both the within-level covariance and residual structures.

When comparing the performance of each fit index across different types of misspecifications (see Figures 36 – 41), it is noticed that all fit indices react similarly to misspecifications in the between-level structure of the model. To be more specific, the distributions of all fit indices to models misspecified in the between-level covariance structure coincide with those when the model is true (i.e., each fit index is highly correlated with itself in detecting these two types of model misspecifications), and the distributions of all fit indices to models misspecified in the marginal mean structure are almost identical to those when the model is misspecified in both the marginal mean and the between-level covariance structures (i.e., their correlations are high). That is to say, none of the fit indices can detect misspecifications concerning the between-level covariance structure of the model. Thus when both the marginal mean structure and the between-level covariance structure are misspecified, they can only detect the misspecification in the marginal mean structure.

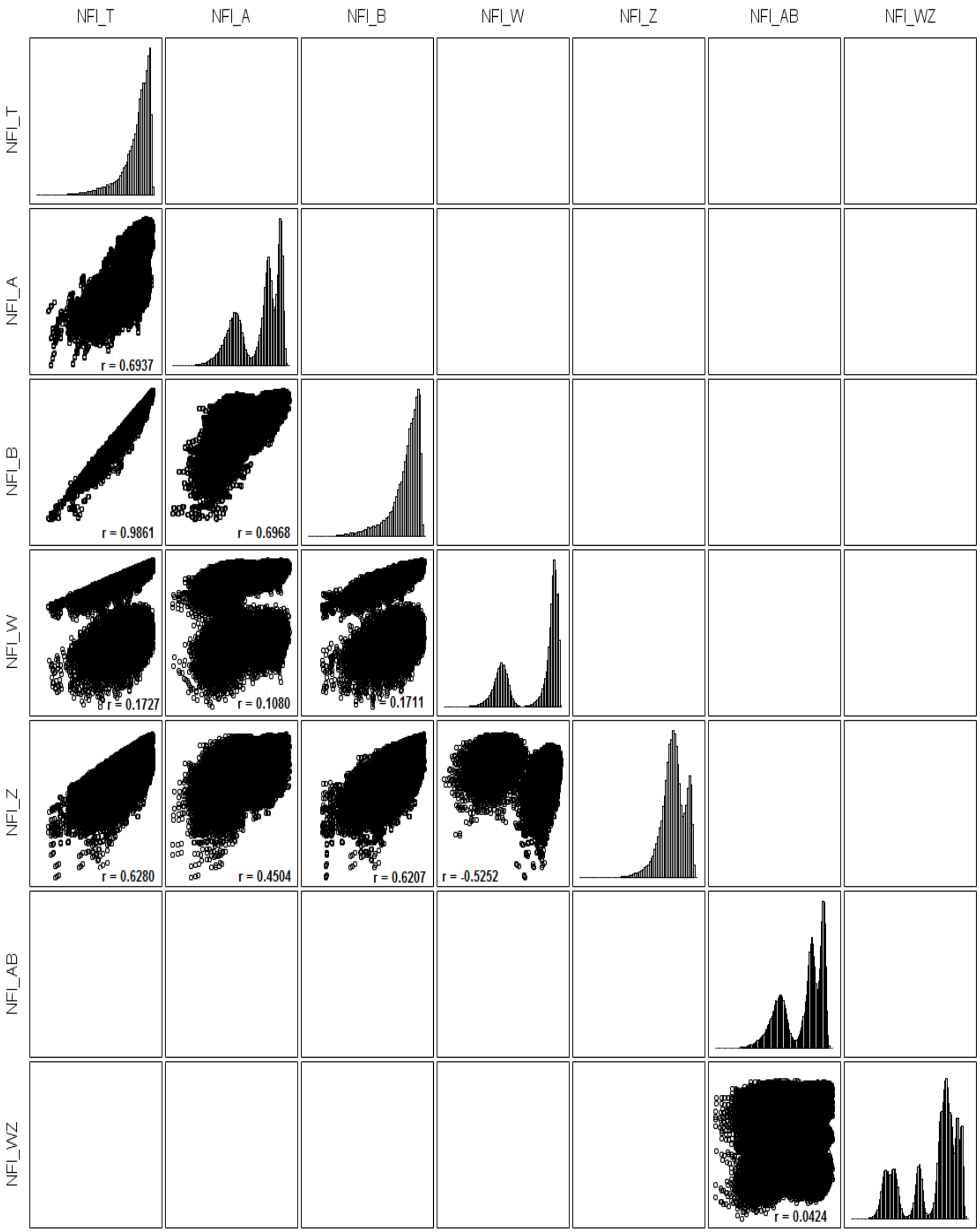


Figure 36: Correlations of *NFI* across types of misspecifications.

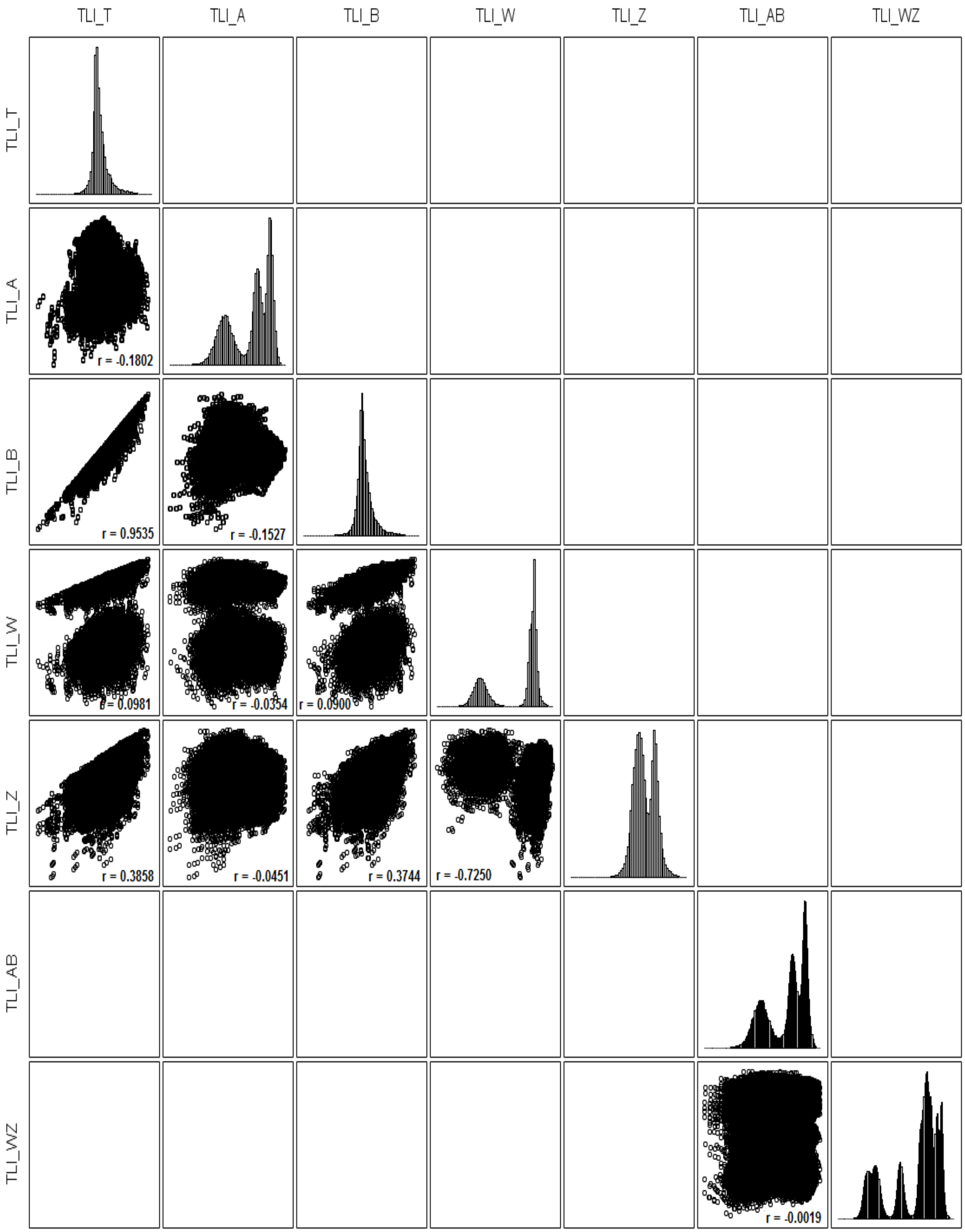


Figure 37: Correlations of *TLI* across types of misspecifications.

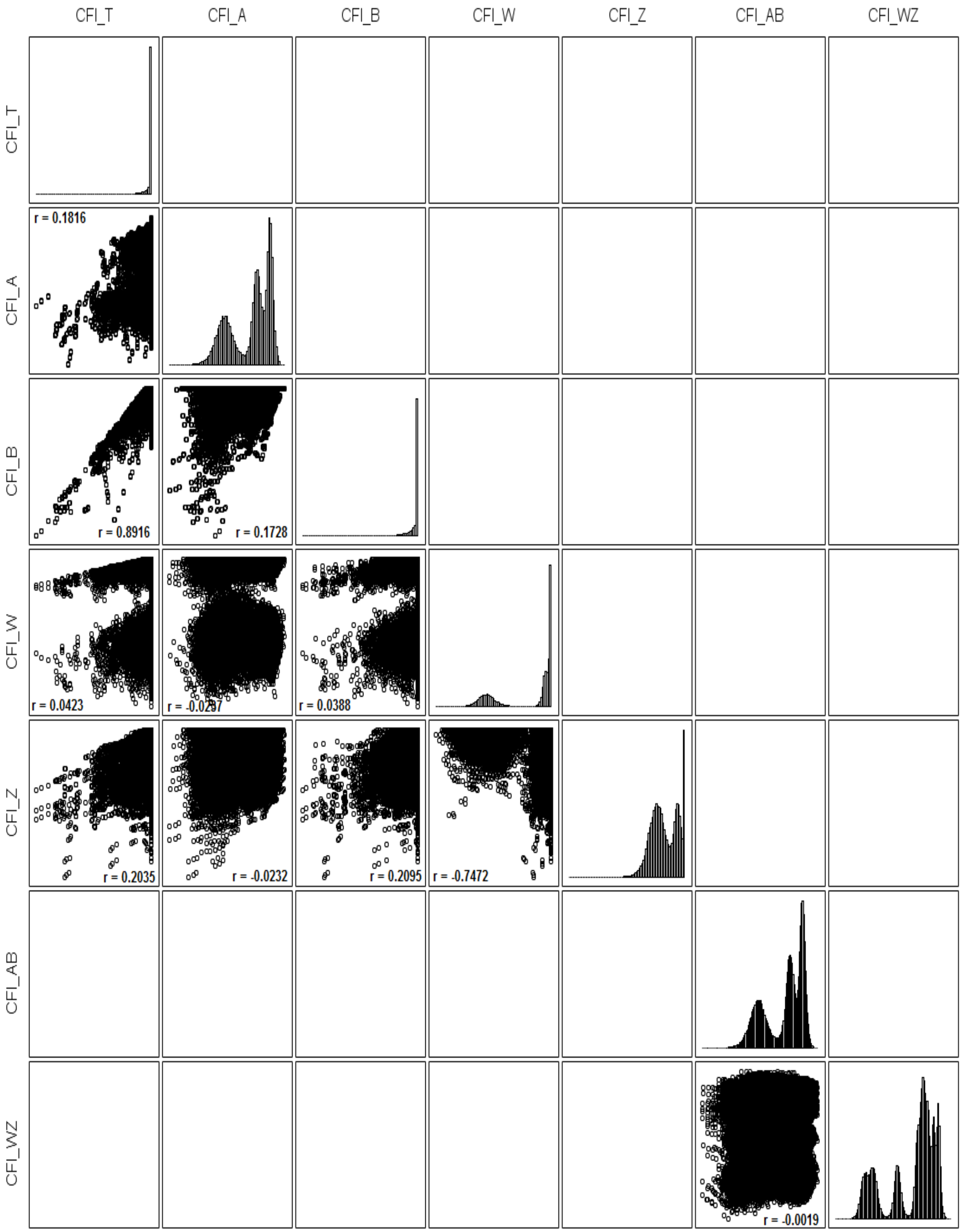


Figure 38: Correlations of *CFI* across types of misspecifications.

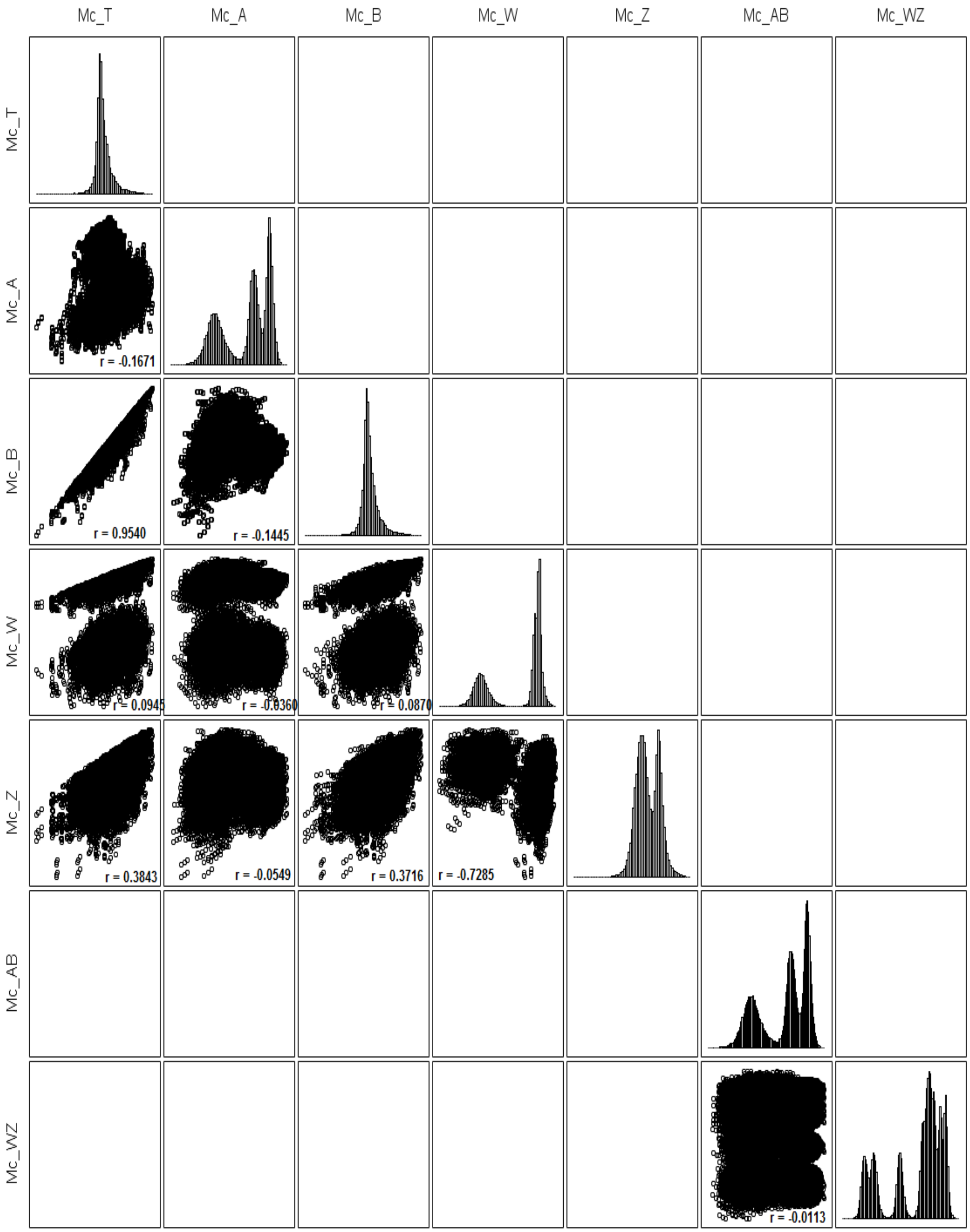


Figure 39: Correlations of *Mc* across types of misspecifications.

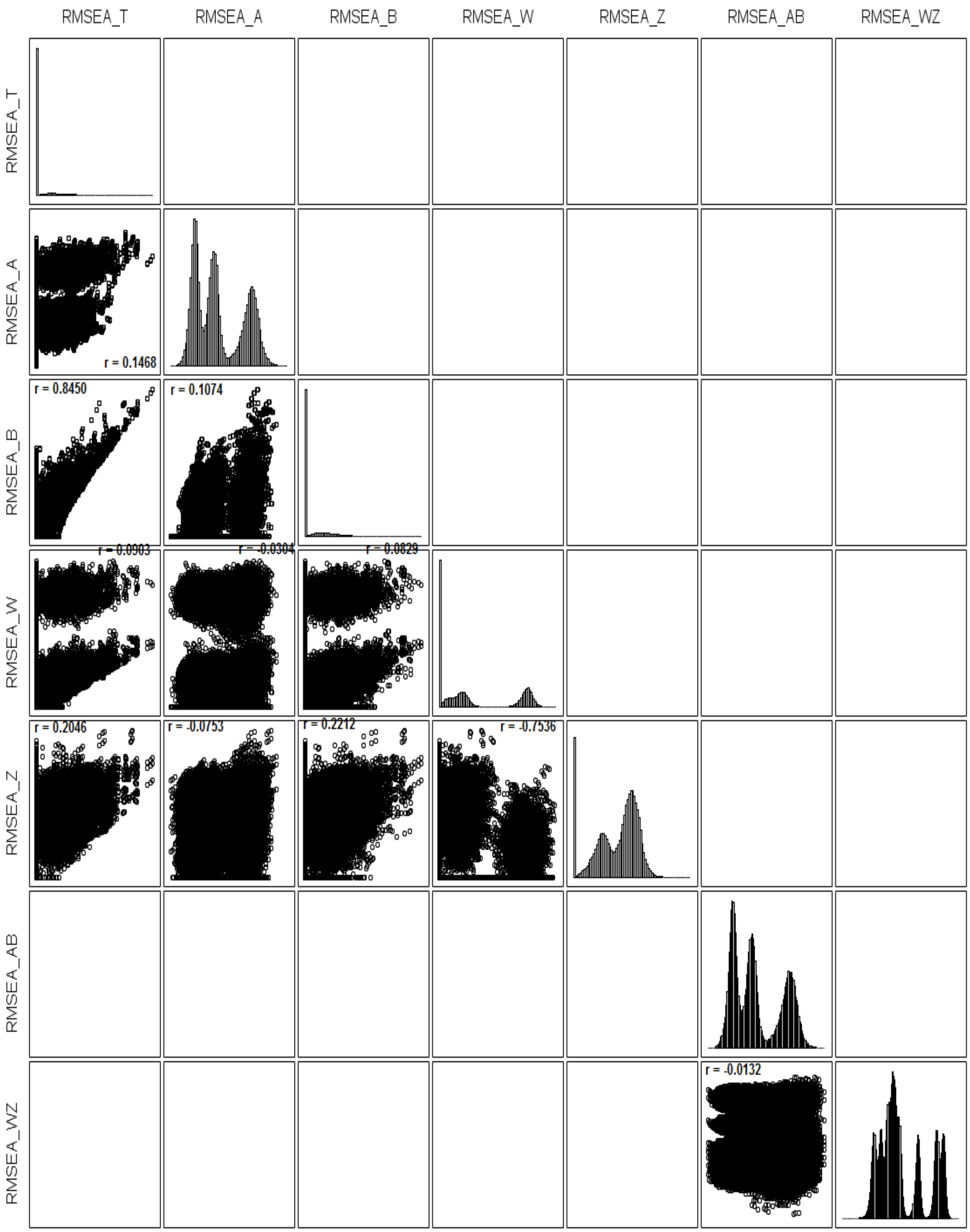


Figure 40: Correlations of *RMSEA* across types of misspecifications.

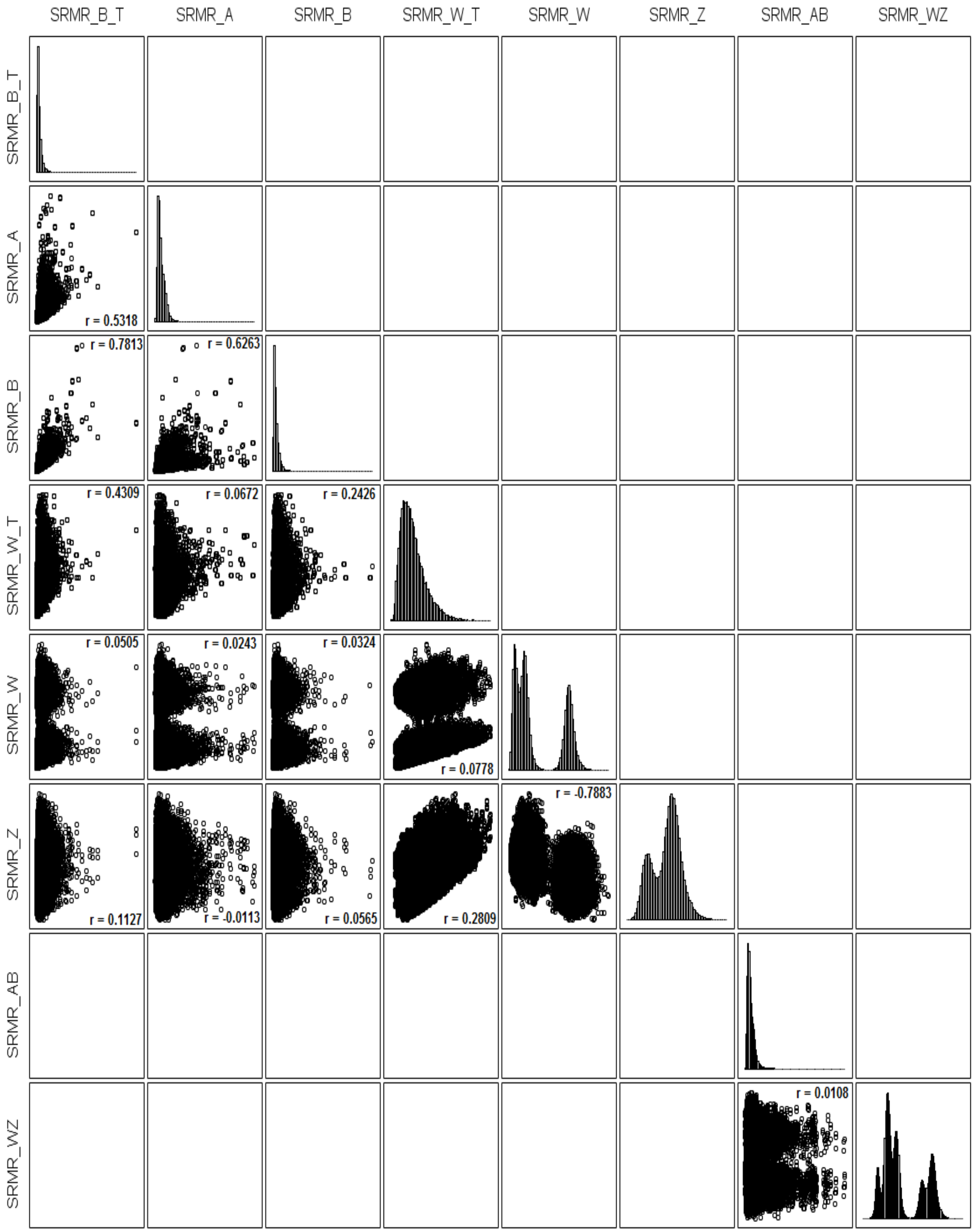


Figure 41: Correlations of *SRMR* across types of misspecifications.

When looking across the correlations among the fit indices and the significant sources of their variability, it is noticed that highly correlated fit indices react to the same sources of variations in a similar way. Low to moderately correlated fit indices, however, are either sensitive to different manipulated factors or responded to the same factors differently.

In sum, *TLI* and *Mc* work interchangeably in detecting misspecifications in linear-linear piecewise multilevel latent growth curve models. Among the six practical fit indices, *SRMR* behaves differentially from other fit indices in that it is majorly sensitive to the intraclass correlation coefficient. None of the fit indices is able to distinguish models misspecified in the between-level covariance structure from the true models, although they are sensitive to misspecifications in the marginal mean structure and those in the within-level structure of the linear-linear piecewise multilevel latent growth curve model.

4.1.5 Cutoff value analyses[34]

Although several cutoff values have been proposed by different authors (e.g., Hu & Bentler, 1999; Yu, 2002), the criteria proposed by Hu and Bentler are used in this study. This is because 1) cutoff values are model and condition based (Hsu, 2009; Yu, 2002), and 2) the criteria suggested by Hu and Bentler is mostly accepted and adopted by practitioners. Type I and Type II error rates are used to evaluate the appropriateness of the recommended cutoff values. As is shown in Table 26, the suggested cutoff values perform very well when the model is true (i.e., the Type I error rates for all fit indices are close to zero). However, they lead to really high Type II error rates when the model is misspecified.

Table 26: Average Type I & II Error Rates of the Fit Indices Adopting the Recommended Cutoff Values

Error	**Model/ Misspecification**	**NFI** (≥ 0.90)	**TLI** (≥ 0.95)	**CFI** (≥ 0.95)	**Mc** (≥ 0.90)	**RMSEA** (≤ 0.06)	**SRMR** (≤ 0.08)
Type I	True	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002[35]
Type II	α_3	1.0000	0.8377	0.9999	0.9970	0.5835	0.9863
	ψ_{B21}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9983
	α_3 & ψ_{B21}	1.0000	0.8792	0.9999	0.9971	0.6227	0.9645
	ψ_{W32}	1.0000	0.9846	1.0000	0.9962	0.6711	1.0000
	ζ_3	1.0000	1.0000	1.0000	1.0000	0.9997	1.0000
	ψ_{W32} & ζ_3	0.7856	0.3210	0.6542	0.3899	0.0965	1.0000

Hence it is hoped that a new set of cutoff values, which controls Type I and Type II errors simultaneously, could be found to detect various types of

[34] Classification trees were also adopted to decide the categorization of different levels of sample size. The results showed that the data were highly skewed and no classification was found.

[35] This is the value based on *SRMR_B*. the Type I error rates for SRMR_W is 0.0000 for the true model.

misspecifications concerning linear-linear piecewise multilevel latent growth curve models. The powers to reject each type of misspecified models on alpha levels of 0.01, 0.05, and 0.10 are computed (see Tables 27 – 29), which demonstrate that the powers to reject models misspecified in both the between-level and the within-level covariance structures decrease with the decrease of Type I error rates.

Table 27: Powers to Detect Misspecified Models with Alpha = 0.10

Misspecification	**NFI**	**TLI**	**CFI**	**Mc**	**RMSEA**	**SRMR**
α_3	1.0000	1.0000	1.0000	1.0000	1.0000	0.8791
ψ_{B21}	0.1111	0.1635	0.1670	0.1717	0.1654	0.3039
α_3 & ψ_{B21}	1.0000	1.0000	1.0000	1.0000	1.0000	0.8921
ψ_{W32}	0.4561	0.6698	0.6718	0.6752	0.6720	0.7482
ζ_3	0.7189	0.9035	0.9043	0.9045	0.9031	0.9250
ψ_{W32} & ζ_3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 28: Powers to Detect Misspecified Models with Alpha = 0.05

Misspecification	**NFI**	**TLI**	**CFI**	**Mc**	**RMSEA**	**SRMR**
α_3	1.0000	1.0000	1.0000	1.0000	1.0000	0.7460
ψ_{B21}	0.0577	0.0794	0.0772	0.0853	0.0806	0.1942
α_3 & ψ_{B21}	1.0000	1.0000	1.0000	1.0000	1.0000	0.7821
ψ_{W32}	0.3935	0.6168	0.6148	0.6232	0.6198	0.7002
ζ_3	0.5270	0.8700	0.8684	0.8716	0.8686	0.8418
ψ_{W32} & ζ_3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 29: Powers to Detect Misspecified Models with Alpha = 0.01

Misspecification	**NFI**	**TLI**	**CFI**	**Mc**	**RMSEA**	**SRMR**
α_3	0.9994	1.0000	1.0000	1.0000	1.0000	0.3910
ψ_{B21}	0.0118	0.0137	0.0141	0.0157	0.0145	0.0552
α_3 & ψ_{B21}	0.9995	1.0000	1.0000	1.0000	1.0000	0.4378
ψ_{W32}	0.3484	0.4706	0.4741	0.4944	0.4858	0.5758
ζ_3	0.1728	0.7114	0.7148	0.7215	0.7128	0.7032
ψ_{W32} & ζ_3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

When the Type I error rate is controlled at 10%, the powers for *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* to detect the misspecification in the between-level covariance structure are no more than 0.2, and the power for *SRMR* to detect that type of misspecification is around 0.3. When the Type I error rates drops to 1%, the powers for the fit indices to detect the within-level covariance structure also decrease to around 0.5. In this sense, it is hard to find a new set of cutoff values that have moderate to high powers to reject all types of misspecifications while controlling the Type I error rates at ordinary levels.

4.2 Results of Study 2

4.2.1 Descriptive statistics

Again replications that do not converge are deleted casewise. Since the population values are the same as those that are adopted in Study 1, the convergence rate is high for the second study as well. The only one replication that does not converge completely is listed below in Table 30.

Table 30: Summary of the Convergence Rates for Models Failing to Achieve Complete Convergence (i.e., 100% Convergence Rate)

Model	**Estimation**	**ICC**	**Severity**	**N_g**	**N_{gi}**	**Convergence Rate (%)**
True	Simultaneous	0.10	0.99	50	10	99.5

Table 31: Descriptive Statistics of the Fit Indices for the True Model

		Simultaneous Estimation				**Partially-Saturated Estimation**			
		True		ψ_{B21}		**True**		ψ_{B21}	
Index	**N**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
NFI	16199	0.9982	0.0014	0.9980	0.0015	0.9914	0.0061	0.9894	0.0074
TLI	16199	1.0008	0.0013	1.0006	0.0013	1.0057	0.0068	1.0038	0.0083
CFI	16199	0.9999	0.0002	0.9999	0.0003	0.9996	0.0013	0.9992	0.0038
Mc	16199	1.0021	0.0032	1.0015	0.0033	1.0347	0.0410	1.0244	0.0455
RMSEA	16199	0.0016	0.0040	0.0028	0.0049	0.0060	0.0171	0.0123	0.0237
SRMR_B	16199	0.0069	0.0061	0.0120	0.0102	0.0069	0.0061	0.0120	0.0106

When partially-saturated estimation method is adopted, the means of all fit indices deviate further away from the ideal values (i.e., the means of *NFI*, *TLI*, *CFI*, and *Mc* become bigger and the means of *RMSEA* and *SRMR* become smaller) when the model is true. When the model is misspecified in the between-level covariance structure, the averages of *NFI* and *CFI* drop and the average of *RMSEA* increases via partially-saturated estimation method when compared to the values obtained via simultaneous estimation method, implying that they have more power to reject the misspecified model via partially-saturated estimation method. However, this improvement in the sensitivity of the three fit indices to misspecified between-level covariance structure is trivial (i.e., the changes in the averages are no more than 0.01), suggesting that partially-saturated estimation method may not contribute much under these data conditions (i.e., small sample sizes and low correlations between the covariance parameters). The averages of *TLI* and *Mc* grow when using the partially-saturated estimation method to detect the misspecification in the between-level covariance structure, indicating that they are more likely to commit Type II errors with this estimation method. The average of *SRMR*, however, does not change, indicating that it is not influenced by the saturation on the other level.

The standard deviations of all fit indices expand when adopting the partially-saturated estimation method to evaluate both the true model and the model misspecified in the between-level covariance structure. This suggests that all fit indices perform less stably when adopting the partially-saturated estimation method.

4.2.2 ANOVA analyses

Nine two-sample Kolmogorov-Smirnov tests[36] are conducted to check the influence of the partially-saturated estimation method on the parameter estimates of the marginal means and the between-level covariance matrices (see Figure 42). The results suggest that this estimation method do not statistically significantly change the parameter values, thus satisfying the independent assumption of linear-linear piecewise multilevel latent growth curve models.

[36] Bias, SE and RMSE may be adopted to calibrate the influence of the partially-saturated estimation method on parameter estimates.

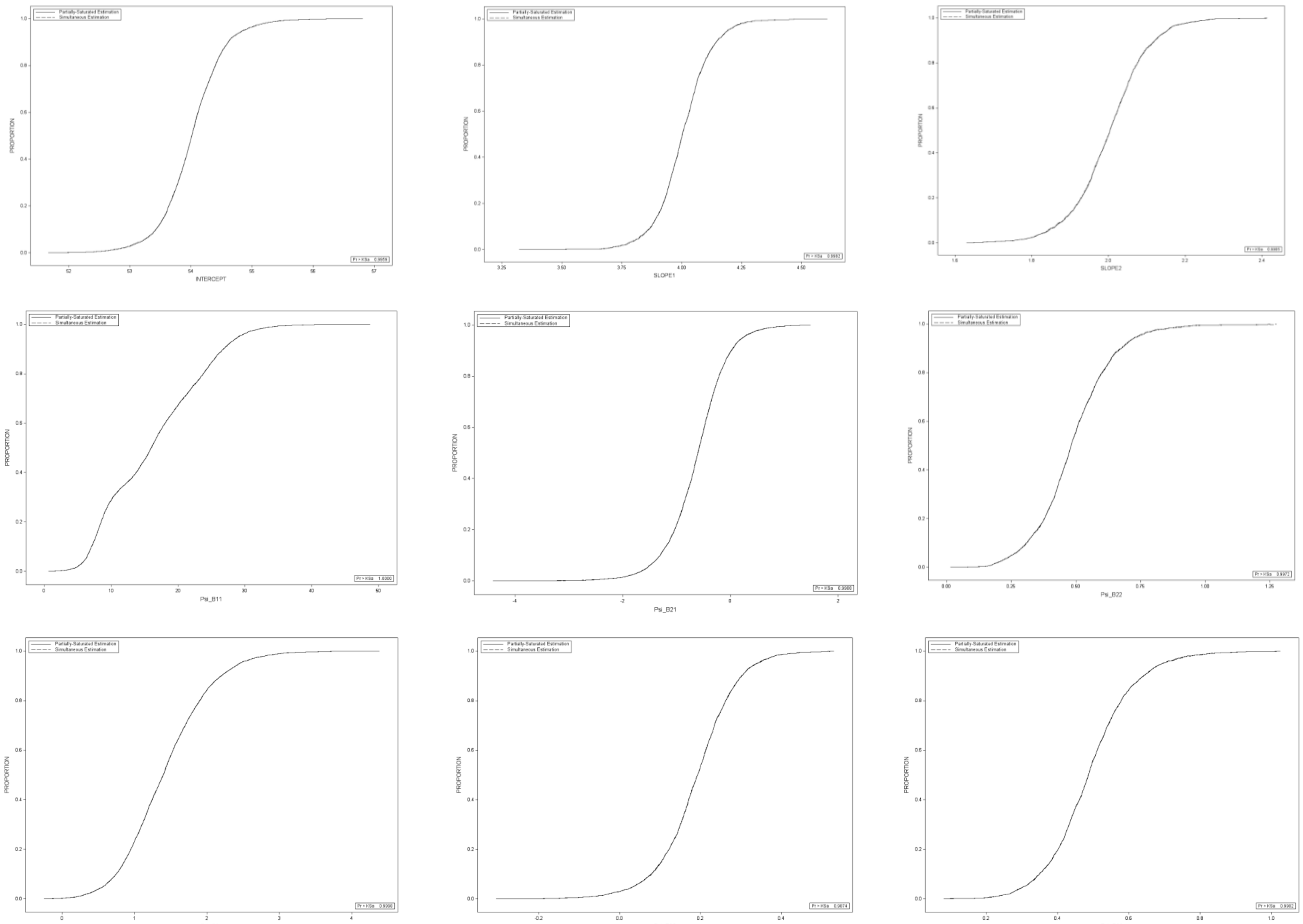


Figure 42: Two-sample Kolmogorov-Smirnov tests for the between-level mean and covariance matrices.

The ANOVA results (see Tables 32 – 33) suggest that the partially-saturated estimation method have a large and significant influence on the variability of *NFI*, *TLI*, and *Mc* when the model is true and when the model is misspecified in the between-level covariance structure. Moreover, it has a moderate but significant influence on *RMSEA* when the misspecification occurs in the between-level covariance structure. However, it does not statistically and practically significantly influence the performance of *CFI* and *SRMR* when the model is either true or misspecified in the between-level covariance structure.

Table 32: Effect Sizes (Partial Eta-Squared ≥0.06) for the Fit Indices for the True Model

Index	**Estimation**	**N_g**	**Estimation×N_g**	**ICC**	**N_{gi}**
NFI	0.3750	0.2206	0.0918		
TLI	0.3399	0.2666	0.1105		
CFI					
Mc	0.2398	0.1141	0.0901		
RMSEA					
SRMR_B		0.1126		0.1893	0.1863

Table 33: Effect Sizes (Partial Eta-Squared ≥0.06) for the Fit Indices for Models Misspecified in the Between-Level Covariance Structure

Index	**Estimation**	**N_g**	**Estimation×N_g**	**ICC**
NFI	0.3871	0.1742	0.0754	
TLI	0.1952	0.2255	0.0894	
CFI				
Mc	0.1120	0.1035	0.0816	
RMSEA	0.0713			
SRMR_B				0.2518

With those fit indices on which the partially-saturated estimation method have a significant influence, this estimation method helps the fit indices be sensitive to the correct sample size (see Tables 34 – 35). In other words, instead of being sensitive to group size when the simultaneous estimation method is adopted, *NFI*, *TLI*, and *Mc* are sensitive to group number when the partially-saturated estimation method is used.

Table 34: Effect Sizes (Partial Eta-Squared ≥0.06) for the Fit Indices for the True Model

	Simultaneous Estimation				**Partially-Saturated Estimation**		
Index	$\mathbf{N_g}$	$\mathbf{N_{gi}}$	$\mathbf{N_g \times N_{gi}}$	**ICC**	$\mathbf{N_g}$	$\mathbf{N_{gi}}$	**ICC**
NFI	0.4193	0.3055	0.0789		0.5043		
TLI	0.1526	0.0886			0.2552		
CFI							
Mc	0.1540	0.0944			0.2694		
RMSEA							
SRMR_B	0.1119	0.1853		0.1882	0.1133	0.1873	0.1904

Table 35: Effect Sizes (Partial Eta-Squared ≥0.06) for the Fit Indices for Models Misspecified in the Between-Level Covariance Structure

	Simultaneous Estimation				**Partially-Saturated Estimation**	
Index	$\mathbf{N_g}$	$\mathbf{N_{gi}}$	$\mathbf{N_g \times N_{gi}}$	**ICC**	$\mathbf{N_g}$	**ICC**
NFI	0.4133	0.3126	0.0778		0.4071	
TLI	0.1422	0.0696			0.1469	
CFI						
Mc	0.1409	0.0730			0.2087	
RMSEA						
SRMR_B				0.2641		0.2404

The influence of group number on *NFI*, *TLI*, and *Mc* when the partially-saturated estimation method is used is illustrated in Figures 43 – 45. Compared to the simultaneous estimation method, partially-saturated estimation method sharpens the influence of group number on *NFI*, *TLI*, and *Mc* in that the averages of those fit indices change more abruptly when the group number increases.

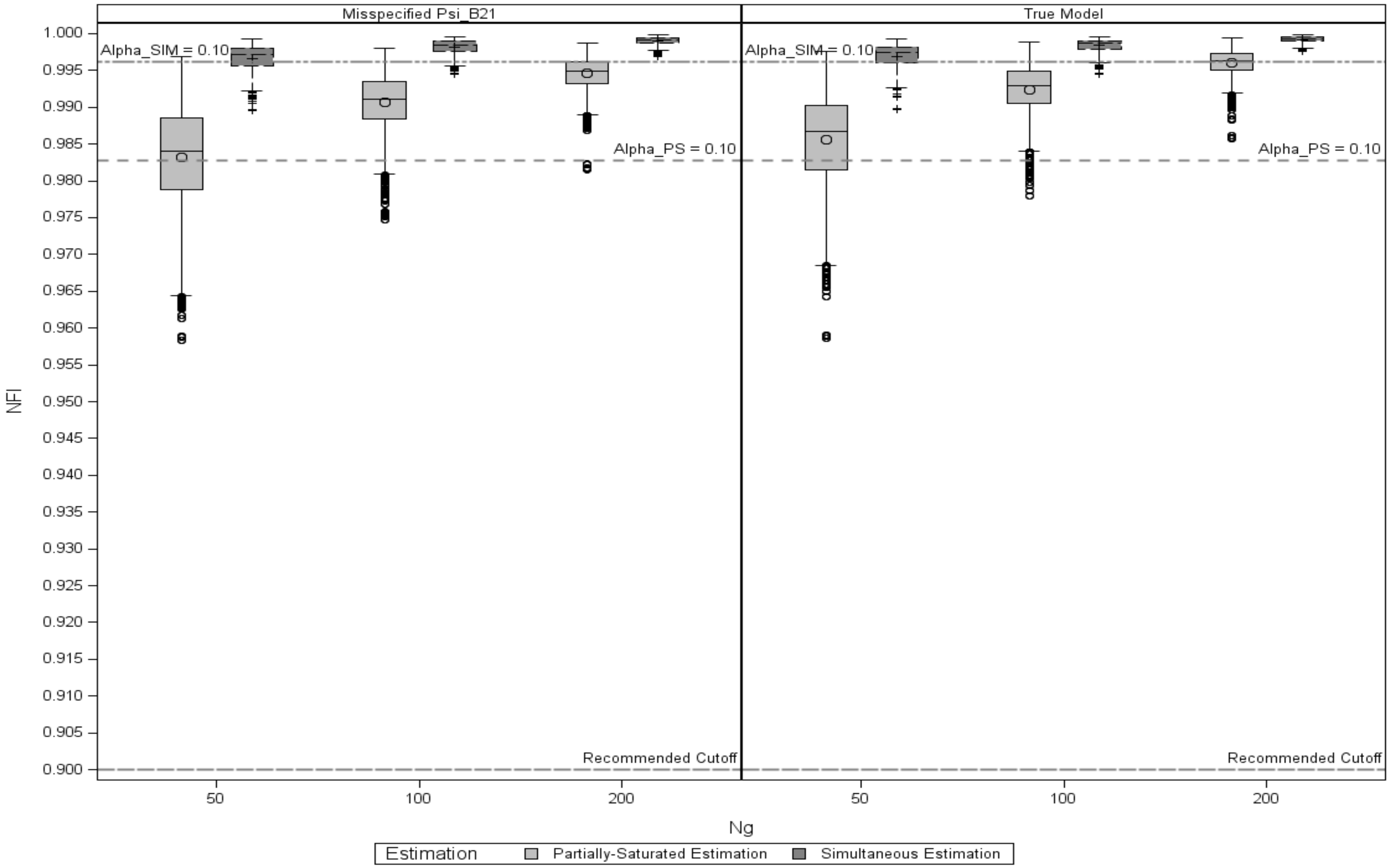


Figure 43: The effects of (estimation method×N_g) on the performance of *NFI*.

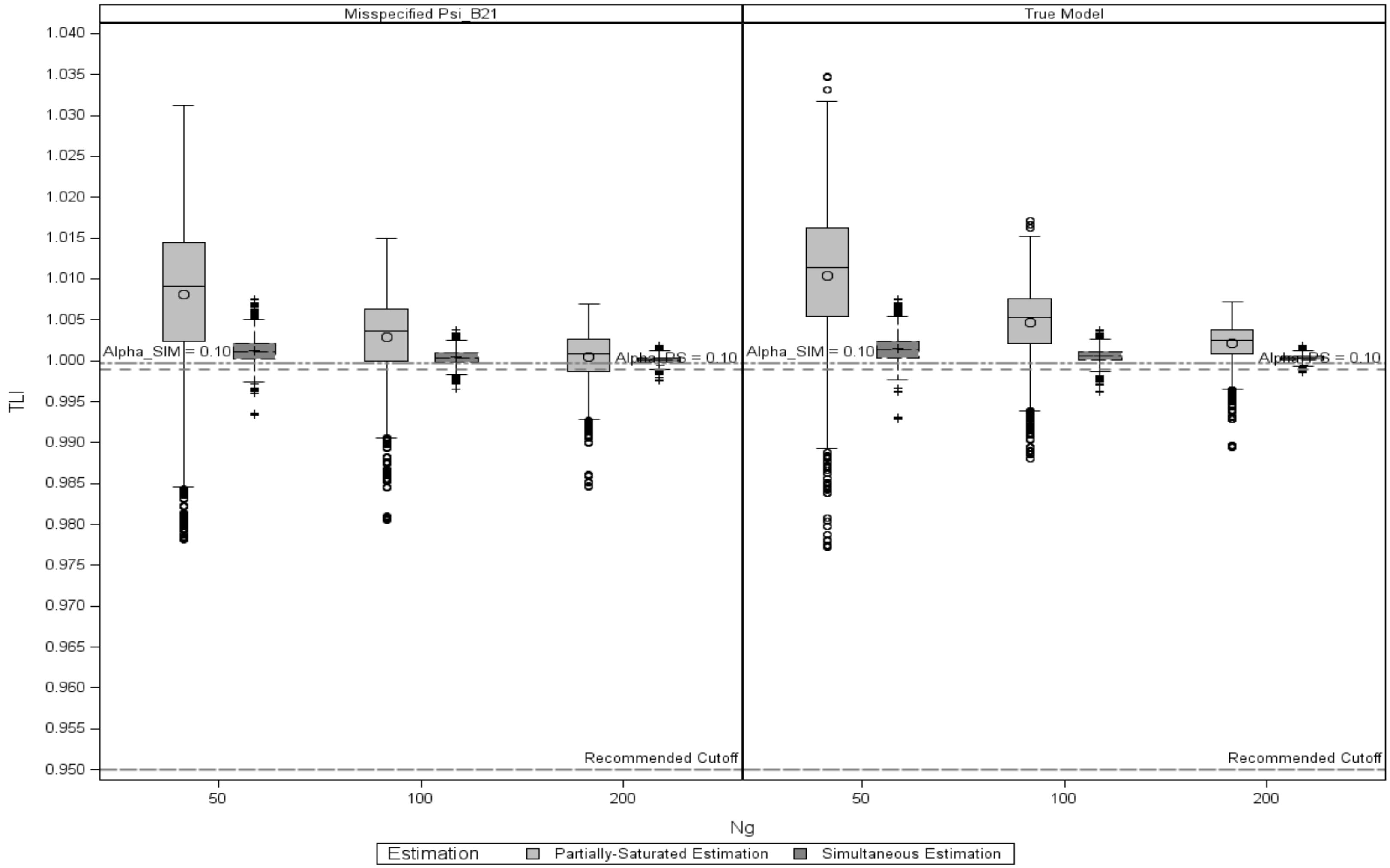


Figure 44: The effects of (estimation method×N_g) on the performance of *TLI*.

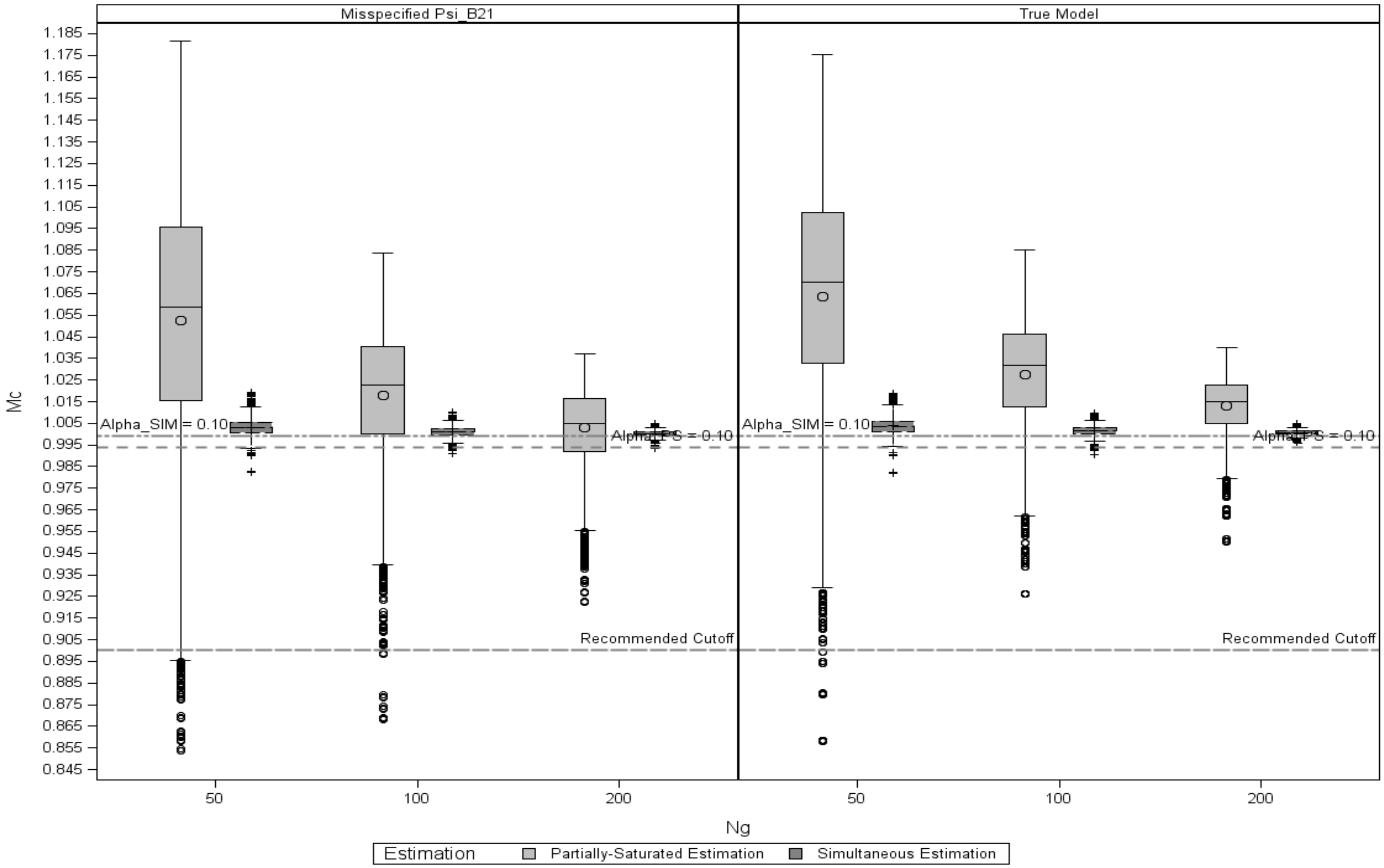


Figure 45: The effects of (estimation method×N_g) on the performance of *Mc*.

4.2.3 Correlation analyses

Although ANOVA results suggest that the partially-saturated estimation method does not practically significantly influence all fit indices, it changes their distributions when detecting misspecification in the between-level covariance structure such that all fit indices respond more similarly to that type of misspecification (see Figure 46). The correlations among all fit indices increase obviously, with *SRMR* still being less correlated with other fit indices. When the model is true, the partially-saturated estimation method expands the spread of *NFI*, *TLI*, and *Mc* (see Figure 47), which consequently increases their associations with other fit indices. *SRMR*, however, is less correlated with other fit indices via the partially-saturated estimation method when the model is true.

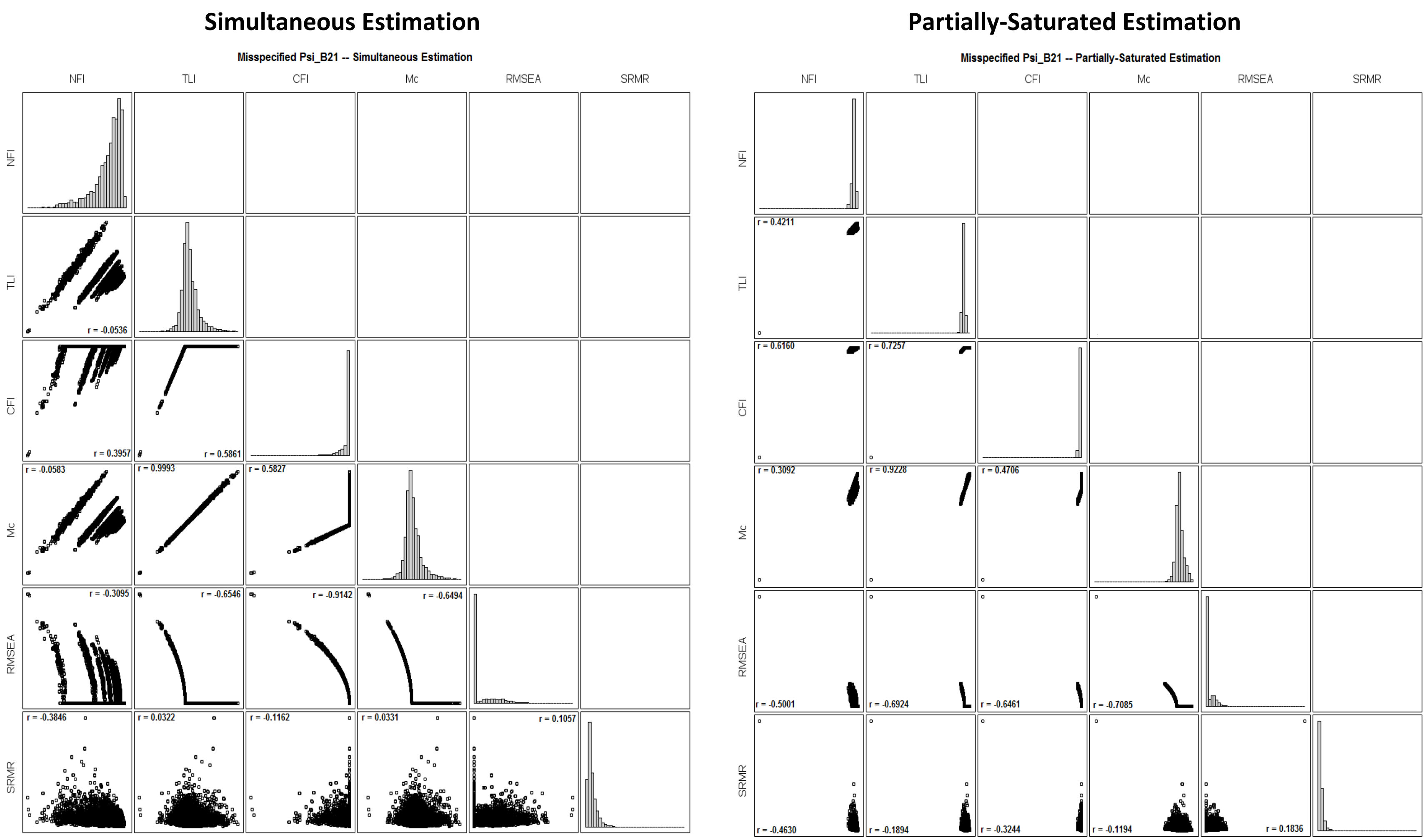


Figure 46: Correlations of the fit indices for models misspecified in the between-level covariance structure.

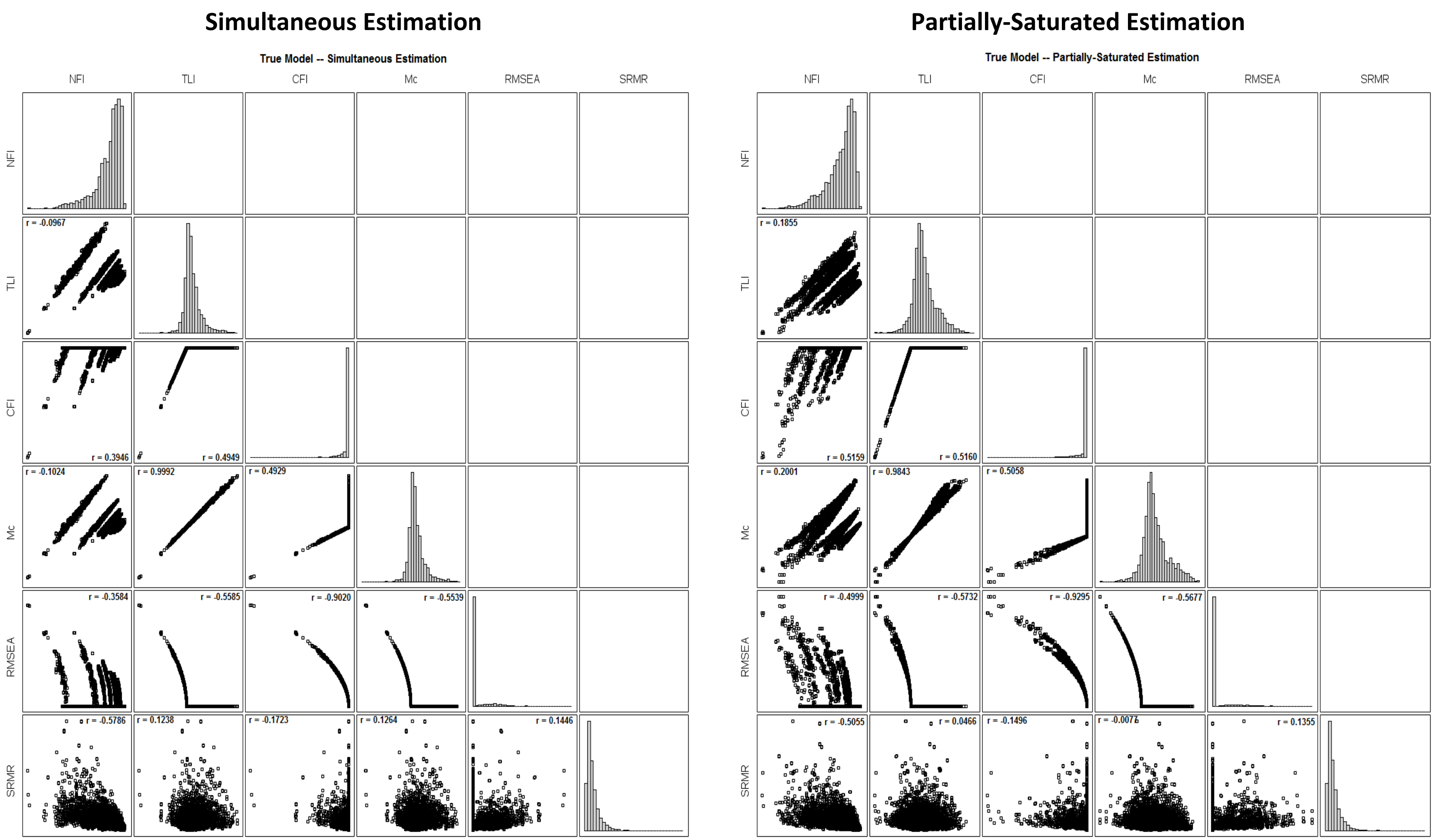


Figure 47: Correlations of the fit indices for the true model.

4.2.4 Cutoff value analyses

The cutoff values suggested by Hu and Bentler (1999) are reevaluated to check whether they work better with the partially-saturated estimation method (see Table 36). Compared to the simultaneous estimation method, the partially-saturated estimation method leads to higher Type I error rates for *Mc* and *RMSEA* when the model is true, but lower Type II error rates for all fit indices when the model is misspecified in the between-level covariance structure. This is because the partially-saturated estimation method contributes to the decrease of the spread for all fit indices.

Table 36: Average Type I & II Error Rates of the Fit Indices Adopting the Recommended Cutoff Values

	Type I Error Rate		**Type II Error Rate**	
Index	**Simultaneous Estimation**	**Partially-Saturated Estimation**	**Simultaneous Estimation**	**Partially-Saturated Estimation**
NFI	0.0000	0.0000	1.0000	0.9999
TLI	0.0000	0.0000	1.0000	0.9999
CFI	0.0000	0.0000	1.0000	0.9999
Mc	0.0000	0.0017	1.0000	0.9951
RMSEA	0.0000	0.0306	1.0000	0.9388
SRMR_B	0.0004	0.0004	0.9981	0.9980

Table 37: Powers to Detect Models Misspecified in the Between-Level Covariance Structure via Different Estimation Methods

	Simultaneous Estimation			**Partially-Saturated Estimation**		
Index	α **= 0.01**	α **= 0.05**	α **= 0.10**	α **= 0.01**	α **= 0.05**	α **= 0.10**
NFI	0.0180	0.0607	0.1223	0.0217	0.0868	0.1567
TLI	0.0144	0.1003	0.1783	0.0216	0.1009	0.2146
CFI	0.0159	0.0916	0.1696	0.0252	0.1070	0.2195
Mc	0.0154	0.0959	0.1723	0.0250	0.1137	0.2153
RMSEA	0.0141	0.0896	0.1746	0.0227	0.1074	0.2120
SRMR_B	0.0412	0.1770	0.2863	0.0413	0.1802	0.2846

The powers for all fit indices to reject models misspecified in the between-level covariance structure on alpha levels of 0.01, 0.05, and 0.10 are recalculated (see Table 37), with the purpose to examine the possibility of finding out new cutoff criteria for the fit indices. However, even with the partially-saturated estimation method, the power for all fit indices to reject models misspecified in the between-level covariance structure are still low when controlling for the Type I error rates (i.e., the powers are no larger than 0.3 across different Type I error rates), suggesting that it is still impossible to set up new criteria for those fit indices which could help practitioners distinguish the misspecified from the true between-level covariance structure in linear-linear piecewise multilevel latent growth curve models.

Chapter 5: Conclusions & Discussion

No study has been conducted to investigate the sensitivity of practical fit indices in detecting misspecifications in linear-linear piecewise multilevel latent growth curve models. This dissertation examines the sensitivity of six practical fit indices in detecting misspecifications concerning the marginal mean structure, the between-level covariance structure, the within-level covariance structure, the within-level residual structure, both the marginal mean and the between-level covariance structures, and both the within-level covariance and residual structures of a linear-linear piecewise multilevel latent growth curve model, based on continuous multivariate normal data. In addition to the complexity of the model, this study adopts small sample sizes on both the group and the individual levels to investigate the performance of the fit indices in detecting misspecifications and the influence of partially-saturated estimation method on the performance of the fit indices. The major findings of this study are listed below:

- All fit indices are more sensitive to misspecifications on the within level structure than those on the between level structure of the model. On the within level of the model, all fit indices are more sensitive to the misspecification in the covariance structure than that in the residual structure; on the between level of the model, all of them are more sensitive to the misspecification in the marginal mean structure than that in the covariance structure. Actually, none of the fit indices are practically significantly sensitive to the misspecification in the between-level covariance structure.

- All fit indices except for *NFI* are highly sensitive to the severity of misfit if the misspecification happens in the within-level covariance structure or in both the within-level covariance and the residual structures. When the misspecification occurs in the residual structure, *NFI* is additionally largely influenced by the group and the individual level sample sizes. In addition, all fit indices are more sensitive to the severity levels of misfit in the within-level covariance structure than those in the within-level residual structure.
- The selected fit indices respond differently to the misspecifications on the between level structure of the model. *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are principally influenced by group size when the misspecification involves either the marginal mean structure or both the marginal mean and the between-level covariance structures. *SRMR*, however, is largely affected by the intraclass correlation coefficient when the above two types of misspecifications occur. When the misspecification involves only the between-level covariance structure, *NFI*, *TLI*, and *Mc* are moderately to largely influenced by both the group and the individual level sample sizes, with the influence from the group level sample size being larger than that from the individual-level sample size. Among the three fit indices, *NFI* is the one that is most severely affected by sample size in that it is

additionally affected by the interactions between the two types of sample size. On the contrary, *CFI* and *RMSEA* are not practically significantly influenced by either type of sample size when the model is misspecified in the between-level covariance structure. In contrast to all other fit indices, *SRMR* is principally and practically significantly sensitive to the intraclass correlation coefficient when the model is misspecified on the between-level structure. It is reasonable that NFI, TLI, CFI, Mc, and RMSEA are sensitive to different types of misspecifications and sample size when they occur on the between-level structure of the model since all these fit indices are chi-square based fit indices. SRMR, however, is a residual-based index and might not easily detect misspecifications on the between-level.

- *NFI*, *TLI*, *CFI*, *Mc*, and *RMSEA* are more likely to commit Type II errors in detecting misspecifications in the marginal mean structure and in both the marginal mean and the covariance structures if group size increases. *SRMR*, however, would commit more Type II errors in detecting misspecification in the covariance structure but less Type II errors if the misspecification involves the marginal mean and the between-level covariance structures. In addition, *TLI* and *Mc* enjoy more power in detecting the misspecification in the between-level covariance structure with the increment of group size.
- When the number of groups increases, *TLI* and *Mc* have more power in detecting the misspecification in the between-level covariance structure whereas *NFI* is more likely to commit Type II errors.

- All fit indices have more power to detect the misspecification in the within-level covariance structure if the severity levels of the misspecification increase; however, they are more likely to commit Type II errors if the misspecification concerns the residual structure. If the misspecification involves both the within-level covariance and residual structures, all fit indices have more power to detect the misspecification within each severity level of the misspecified residual structure.
- When the intraclass correlation coefficient increases, *SRMR* is more likely to commit Type II errors if the model is misspecified in the between-level structure.
- *Mc* may function as a substitute for *TLI* in detecting misspecifications in the linear-linear piecewise multilevel latent growth curve model. *SRMR* behaves differentially from all other fit indices when detecting misspecifications involving the between-level structure of the model.
- Partially-saturated estimation method has a moderate to large influence on the performance of *NFI*, *TLI*, *Mc*, and *RMSEA* when detecting the misspecification in the between-level covariance structure. Moreover, it helps *NFI*, *TLI*, and *Mc* to be sensitive to the appropriate sample size (i.e., group number) and sharpens the effect of that sample size on the performance of the three fit indices. However, this estimation method does not have a statistically and practically significant influence on the performance of *CFI* and *SRMR*.

- Even with the fit indices that the partially-saturated estimation method has a significant influence (i.e., *NFI*, *TLI*, *Mc*, and *RMSEA*) in detecting the misspecification in the between-level covariance structure, this estimation method does not help them to be sensitive to the severity level of the misfit.
- The suggested cutoff value for each fit index controls for Type I errors very well; however, they result in really high Type II errors across all types of misspecifications if simultaneous estimation method is adopted. In addition, they lead to low powers to reject models misspecified in between-level covariance structure when controlling for Type I error rate via the partially-saturated estimation method.
- Given that none of the fit indices are really sensitive to the misspecification in the between-level covariance via both simultaneous and partially-saturated estimation methods, it is almost infeasible to find a new set of cutoff values which can control Type I and Type II errors simultaneously.

The above results suggest that all six fit indices could be used to evaluate the appropriateness of a linear-linear piecewise latent growth curve model, even with small sample size. However, those fit indices could not tell practitioners which kind of misspecification a model may involve. Instead, practitioners could adopt modification indices to revise the model step by step.

Given the fact that the within-level sample size is much larger than the between-level sample size, it is quite natural to observe a large and significant influence of group size on the performance of *NFI, TLI, CFI, Mc* and *RMSEA* when detecting the misspecification in the marginal mean structure. *SRMR* is not practically significantly influenced by either type of sample size because it is a residual-based index, thus removing the influence of sample size.

A noticeable finding of this study is that none of the fit indices is sensitive to the misspecification in the between-level covariance structure, no matter which estimation method is used. One possible explanation for this phenomenon is that the group level sample size adopted in this study is far too small, which hinders the demonstration of the asymptotic feature associated with chi-square test; another possible reason is that the magnitude of the covariance between the intercept and the first-phase slope parameters is too small (i.e., the correlation between the two parameters is around 0.2), making any changes in the severity level barely detectable (i.e., the correlations corresponding to different severity levels are bounded within the range of 0.1 to 0.2, all of which are small and ignorable). Further studies may increase the magnitude of the correlations among the parameters in the between-level covariance structure and re-investigate the performance of the fit indices.

In addition to the above potential improvement in the research design, this study may be extended in other aspects to make it more complete and informative.

- This study employs an ideal condition (i.e., multivariate normally distributed outcomes measured at fixed time points and without missing data) to investigate the sensitivity of the practical fit indices. Future study could include categorical data or continuous but non-normally distributed outcomes, with or without missing data, to examine the performance of the fit indices.

- A linear-linear piecewise multilevel latent growth curve model is examined in this study, which however could be extended to cover other functional forms such as quadratic, exponential, and logistic growth curves to enrich the result information.
- The knot position and the number of observations included in each time segment which are believed to influence the convergence rates and the estimates of parameters and their standard errors could be manipulated to their influence on the performance of the fit indices.
- The normality assumption of the distributions of the fit indices should be checked in advance before conducting the ANOVA analyses. If the normality assumption is not satisfied, data transformation could be implemented beforehand.
- This study controls both the Type I and Type II errors simultaneously to find the suggested cutoff values. In addition to this method, bootstrapping could be adopted to find the empirical distributions of the fit indices, based on which proper cutoff values might be found.
- It is expected that higher ICC values lead to higher SRMR_B values for misspecified between-level covariance structure. However, the reversed correlations are found in this study, whose reasons are unclear and may requests for further exploration.

Appendix

Let S2' be the deviation of slope one (i.e., S1) from slope two (i.e., S2) after the time knot (e.g., time point a), then

$$S_2^{'} = S_2 - S_1 \quad (t \geq a)$$

Because $\begin{pmatrix} I \\ S_1 \\ S_2 \end{pmatrix} \sim MVN\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, S_2' is a normal random variable such that

$\begin{pmatrix} I \\ S_1 \\ S_2^{'} \end{pmatrix} \sim MVN\left(\boldsymbol{\mu}^{'}, \boldsymbol{\Sigma}^{'}\right)$ (Theorem 5.2, Hardle & Hlavka, 2007). Therefore,

$$E\left(S_2^{'}\right) = E\left(S_2 - S_1\right) = E\left(S_2\right) - E\left(S_1\right) = \mu_{S_2} - \mu_{S_1}$$

$$\begin{aligned}
Var\left(S_2^{'}\right) &= Var\left(S_2 - S_1\right) \\
&= E\left[\left(S_2 - S_1\right) - E\left(S_2 - S_1\right)\right]^2 \\
&= E\left\{\left[S_2 - E\left(S_2\right)\right] - \left[S_1 - E\left(S_1\right)\right]\right\}^2 \\
&= E\left\{\left[S_2 - E\left(S_2\right)\right]^2 - 2\left[S_2 - E\left(S_2\right)\right]\left[S_1 - E\left(S_1\right)\right] + \left[S_1 - E\left(S_1\right)\right]^2\right\} \\
&= E\left[S_2 - E\left(S_2\right)\right]^2 - 2E\left[S_2 - E\left(S_2\right)\right]\left[S_1 - E\left(S_1\right)\right] + E\left[S_1 - E\left(S_1\right)\right]^2 \\
&= Var\left(S_2\right) - 2Cov\left(S_2, S_1\right) + Var\left(S_1\right)
\end{aligned}$$

$$\begin{aligned}
Cov\left(I, S_2^{'}\right) &= Cov\left[I, \left(S_2 - S_1\right)\right] \\
&= E\left[I - E\left(I\right)\right]\left[\left(S_2 - S_1\right) - E\left(S_2 - S_1\right)\right] \\
&= E\left[I - E\left(I\right)\right]\left\{\left[S_2 - E\left(S_2\right)\right] - \left[S_1 - E\left(S_1\right)\right]\right\} \\
&= E\left[I - E\left(I\right)\right]\left[S_2 - E\left(S_2\right)\right] - E\left[I - E\left(I\right)\right]\left[S_1 - E\left(S_1\right)\right] \\
&= Cov\left(I, S_2\right) - Cov\left(I, S_1\right)
\end{aligned}$$

$$
\begin{aligned}
Cov\left(S_1, S_2^{'}\right) &= Cov\left[S_1, \left(S_2 - S_1\right)\right] \\
&= E\left[S_1 - E\left(S_1\right)\right]\left[\left(S_2 - S_1\right) - E\left(S_2 - S_1\right)\right] \\
&= E\left[S_1 - E\left(S_1\right)\right]\left\{\left[S_2 - E\left(S_2\right)\right] - \left[S_1 - E\left(S_1\right)\right]\right\} \\
&= E\left[S_1 - E\left(S_1\right)\right]\left[S_2 - E\left(S_2\right)\right] - E\left[S_1 - E\left(S_1\right)\right]\left[S_1 - E\left(S_1\right)\right] \\
&= E\left[S_2 - E\left(S_2\right)\right]\left[S_1 - E\left(S_1\right)\right] - E\left[S_1 - E\left(S_1\right)\right]^2 \\
&= Cov\left(S_2, S_1\right) - Var\left(S_1\right)
\end{aligned}
$$

Bibliography

Alker, H. R. (1969). A typology of ecological fallacies. In M. Dogan & S. Rokkam (Eds.), *Social ecology* (pp. 69–86). Boston, MA: MIT Press.

Anderson, J. C & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155–173.

Anumendem, D. N. (2011). *Advanced techniques in multilevel growth curve modeling: Application to educational effectiveness research*. Unpublished doctoral dissertation, University of Leuven.

Asparouhov, T. & Muthén, B. O. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. In *Proceedings of the American Statistical Association* (pp. 2531–2535). Alexandria, VA: American Statistical Association.

Barr, C. D. (2008). A multilevel modeling alternative to aggregating data in 360 degree feedback. Poster session presented at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 134–167.

Bearden, W. O., Subhash, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*(4), 425–430.

Beauducel, A. & Wittmann, W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*(1), 41–75.

Ben-Shlomo, Y., McCarthy, A., Hughes, R., Tilling, K., Davies, D., & Smith, G. D. (2008). Immediate postnatal growth is associated with blood pressure in young adulthood: The Barry Caerphilly Growth Study. *Hypertension, 52*, 638–644.

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107,* 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* New York: Guilford Press.

Blalock, H. M. (1979). The presidential address: Measurement and conceptualization problems: The major obstacle to integrating theory and research. *American Sociological Review, 44,* 881–894.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W.

Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco: Jossey-Bass.

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

Bollen, K. A. & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Breivik, E. & Olsson, U. H. (2001). Adding variables to improve fit: The effect of model size on fit assessment in LISREL. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 169–194). Chicago: Scientific Software International.

Browne, M. W. (1993). Structural latent curve models. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future direction 2* (pp. 171–197). New York: Elsevier.

Browne, M. W. & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In G. A. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York: Plenum Press.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Browne, M. W. & du Toit, S. H. C. (1991). Models for learning data. In L. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47–68). Washington, DC: American Psychological Association Press.

Browne, W. (2006). Sample size calculation in multilevel modeling. Paper presented at the ESRC Research Methods Festival, Oxford.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.

Busing, F. (1993). *Distribution characteristics of variance estimates in two-level models*. Unpublished manuscript, Leiden University, the Netherlands.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research, 36*(4), 462–494.

Cheong, J. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling*, *18*, 195–211.

Cheung, M. W.-L. & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, *12*(4), 598–619.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (2000). A two-stage approach to multilevel structural equation models: Application to longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples* (pp. 33–50). Mahwah, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Erlbaum.

Crawford, A. M., Pentz, M. A., Chou, C., Li, C., & Dwyer, J. H. (2003). Parallel developmental trajectories of sensation seeking and regular substance use in adolescents. *Psychology of Addictive Behaviors*, *17*, 179–192.

Cronbach, L. J. (1976). Research on classrooms and schools: Formulation of questions, design and analysis. Occasional paper of the Stanford Evaluation Consortium, Stanford University.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*, 371–403.

Cudeck, R. & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*, 41–63.

Curran, P. J. (1994). *The robustness of confirmatory factor analysis to model misspecification and violations of normality.* Unpublished doctoral dissertation, Arizona State University.

Curran, P. J. (2000). A latent curve framework for studying developmental trajectories of adolescent substance use. In J. Rose, L. Chassin, C. Presson, & J. Sherman (Eds.), *Multivariate Applications in Substance Use Research*, (pp. 1–42). Hillsdale, NJ: Erlbaum.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*, 529–569.

Curran, P. J. & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*(4), 526–544.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognitive Development*, *11*(2), 121–136.

Davey, A., Savla, J., & Luo, Z. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling, 12,* 578–597.

Diallo, T. M. O. & Morin, A. J. S. (2014). Power of Latent Growth Curve Models to Detect Piecewise Linear Trajectories. *Structural Equation Modeling*, 00, 1–12.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analyses. *American Journal of Public Health*, *88,* 216–222.

Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger,H. (2007). Challenges in nonlinear structural equation modeling. *Methodology*, *3,* 100–114.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, *2*(2), 119−144.

Duncan, T. E., Alpert, A., & Duncan, S. C. (1998). Multilevel covariance structure analysis of sibling antisocial behavior. *Structural Equation Modeling, 5,* 211–228.

Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller, M., & Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research*, *32*(3), 275–318.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Duncan, T. E., Duncan, S. C., Strycker, A. L., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. New Jersey, Mahwah: Lawrence Erlbaum Associates, Inc.

Enders, C. K & Finney, S. (2003). SEM fit index criteria re-examined: An investigation of ML and robust fit indices in complex models. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Fan, X., & Fan, X. (2005). Power of latent growth modeling for detecting linear growth: Number of measurements and comparison with other analytical approaches. *The Journal of Experimental Education*, *73*, 121–139.

Fan, X. & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*(3), 343–367.

Fan, X. & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529.

Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on Structural Equation Modeling fit indices. *Structural Equation Modeling*, *6*(1), 56–83.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, *43*, 557–572.

Flora, D. B. (2008). Specifying piecewise latent trajectory models for longitudinal data. *Structural Equation Modeling*, *15*(3), 513–533.

Garst, H. (2000). Longitudinal research using structural equation modeling applied in studies of determinants of psychological well-being and personal initiative in East Germany after the unification. Unpublished doctoral dissertation, University of Amsterdam.

Gerbing, D. W. & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

Goldstein, H. I. (1995). *Multilevel statistical models* (2nd ed.). London: Edwards Arnold.

Gulliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies. *American Journal of Epidemiology*, *149*, 876–883.

Härdle, W. K. & Hlávka, Z. (2007). *Multivariate Statistics: Exercises and Solutions*. Springer.

Harring, J. R., Cudeck, R., & du Toit, S. H. C. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research*, *41*, 579–596.

Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New development and techniques in structural equation modeling* (pp. 89–127). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Heck, R. H. & Thomas, S. L. (2008). *An introduction to multilevel modeling techniques* (2nd ed.). New York, NY: Routledge.

Hedeker, D. & Gibbons, R. D. (2006). *Longitudinal data analysis.* Hoboken, NJ: Wiley.

Hindman, A. H., Cromley, J. G., Skibbe, L. E., & Miller, A. L. (2011). Conventional and piecewise growth modeling techniques: Applications and implications for investigating Head Start children's early literacy leaning. *Evaluation Review*, *35*(3), 204-239.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23,* 723–744.

Hofmann, D. A. (2002). Issues in multilevel research: Theory development, measurement, and analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 247–274). Malden, MA: Blackwell.

Howe, L. D., Tilling, K., Galobardes, B., Smith, G. D., Gunnell, D., & Lawlor, D. A. (2012). Socioeconomic differences in childhood growth trajectories: At what age do height inequalities emerge? *Journal of Epidemiol Community Health*, *66*, 143–148.

Howe, L. D., Tilling, K., Matijasevich, A., Petherick, E. S., Santos, A. C., Fairley, L., Wright, J., Santos, I. S., Barros, A. J. D., Martin, R. M., Krammer, M. S., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. A. (2013). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research, 0*(0), 1–21.

Hox, J. J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In J. H. L. Oud & R. A. W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 141–156). Nijmegen, the Netherlands: ITS.

Hox, J. J. (1998). Multilevel modeling: When and why? In B. R. Mathar & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147–154). New York: Springer Verlag.

Hox, J. J. (2000). Multilevel multivariate and structural equation models: Some missing links. Paper presented at Social Science Methodology in the New Milenium, 5th International Conference on Logic and Methodology, Cologne.

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

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

Hox, J. J. & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*(2), 157–174.

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica, 64*, 157–170.

Hox, J. J. & Stoel, R. D. (2005). Multilevel and SEM approaches to growth curve modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1296–1305). Chichester, UK: Wiley.

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.

Hu, L. & Bentler, P. M. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 76−99). Newbury Park, CA: Sage.

Hu, L. & Bentler, P. M. (1998). Fit indexes in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424−453.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248.

Iacobucci, D. (2009). Structural equation modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*, 90–98.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67,* 219–229.

Jöreskog, K.G. & Sörbom, D. (1981). *LISREL User's guide*. Chicago, National Educational Services.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, *8*, 325–352.

Kaplan, D. & Elliott, P. R. (1997a). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling*, *4*, 1–24.

Kaplan, D. & Elliott, P. R. (1997b). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics*, *22*, 323–347.

Kaplan, D. & George, R. (1998). Evaluating latent variable growth models through ex post simulation. *Journal of Educational and Behavioral Statistics*, *23*(3), 216–235.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*(1), 126.

Kenny, D. A. & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*, 333–351.

Klein, K. J. & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*, 211–236.

Kline, R. B. (2011). Convergence of structural equation modeling and multilevel modeling. In M. Williams & W. P. Vogt (Eds.), *Handbook of methodological innovation in social research methods* (pp. 562–589). London: Sage.

Kline, R. B. (2012). Assumptions of structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 111–125). New York: Guilford Press.

Kohli, N. & Harring, J. R. (2013). Modeling growth in latent variables using a piecewise function. *Multivariate Behavioral Research*, *48*(3), 370–397.

Kohli, N., Harring, J. R. & Hancock, G. R. (2013). Piecewise linear–linear latent growth mixture models with unknown knots. *Educational and Psychological Measurement*, *73*(6), 935–955.

Krull, J. L. & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23,* 418–444.

Krull, J. L. & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277.

Kwok, O., Luo, W., & West, S. G. (2010). Using modification indices to detect turning points in longitudinal data: A Monte Carlo study. *Structural Equation Modeling*. *17*, 216–240.

Kwok, O.-M., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research*, *42,* 557–592.

La Du, T. J. & Tanaka, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology*, 74, 625–635.

Leite, W. L., & Stapleton, L. M. (2011). Detecting growth shape misspecifications in latent growth models: An evaluation of fit indices. *Journal of Experimental Education*, *79*, 361−381.

Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika*, *69*, 101–122.

Little, R. J. A. & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research, 18,* 292–326.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data (2nd ed.)*. New York: Wiley.

Liu, Y., Liu, H., Li, H., & Zhao, Q. (2015). The effects of individually varying times of observations on growth parameter estimations in piecewise growth model. *Journal of Applied Statistics*, 1–18.

Long, S. C. (1984). *Confirmatory factor analysis.* London: Sage.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. O. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203–229.

Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–49.

Markianidou, P. & Weeren, A. (2011). Linear and non linear growth models using mixed modeling: An application on European import volumes. Working Papers 2011-011, University of Antwerp, Faculty of Applied Economics.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391–410.

Marsh, H. W., Balla, J. R., & Hau, K.-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 315–353). Mahwah, NJ: Lawrence Erlbaum Associates.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). New York: Plenum.

McArdle, J. J. (2009). Latent variable modeling of differences in changes with longitudinal data. *Annual Review of Psychology, 60*, 577–605.

McArdle, J. J. & Bell, R. Q. (2000). An introduction to latent growth models for developmental data analysis. In T. D. Little & K. U. Schnabel (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 69–107). Erlbaum: Mahwah, NJ.

McArdle, J. J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110–133.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*, 97–103.

Mehta, P. D., Neale, M. C., & Flay, B. R. (2004). Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes. *Psychological Methods*, *9*, 301–333.

Mehta, P. D. & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*(1), 23–43.

Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research*, *22*, 399–413.

McDonald, R. P. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, *42*, 215–232.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557–585.

Muthén, B. O. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting, Princeton, New Jersey.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 338–354.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398.

Muthén, B. O. (1997a). Latent variable growth modeling with multilevel data. In M. Berkane (Ed.), *Latent variable modeling with applications to causality* (pp. 149–161). Cambridge, MA: Springer-Verlag.

Muthén, B. O. (1997b). Latent variable modeling of longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological methodology* (pp. 453–480). Boston: Blackwell.

Muthén, B. O. (2000). Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking. In J. Rose, L. Chassin, C. Presson, & J. Sherman (Eds.), *Multivariate applications in substance use research* (pp. 113–140). Hillsdale, N.J.: Erlbaum.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, B. O. & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B. O. & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). New York & Hove: Routledge.

Muthén, B. O. & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*, 371–402.

Muthén, B. O., Khoo, S.-T., & Gustafsson, J.-E. (1997). Multilevel latent variable modeling in multiple populations. Retrieved from http://www.statmodel.com/bmuthen/articles/Article_074.pdf.

Muthén, B. O. & Khoo, S.-T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences*, *10*(2), 73–101.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (version 7). Los Angeles, CA: Muthén & Muthén.

Muthén, B. O. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), *Sociological Methodology* (pp. 267–316). Washington, DC: American Sociological Association.

Palardy, G. J. (2003). *A comparison of hierarchical linear and multilevel structural equation growth models and their application in school effectiveness research*. Unpublished doctoral dissertation, University of California, Santa Barbara.

Preacher, K. J. & Hancock, G. R. (2010). Interpretable reparameterizations of growth curve models. Paper presented at the 2010 Advances in Longitudinal Methods conference, College Park, MD. Retrieved from http://www.cilvr.umd.edu/Conference2010/Downloads2010.html

Preacher, K. J., Wichman, A. L., MacCallum, R., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233.

Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 35–64). Washington, DC: The American Psychological Association.

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

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Rogosa, D. R. & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203–228.

Rovine, M. J. & Molenaar, P. C. M. (1998). A nonstandard method for estimating a linear growth model in *LISREL*. *International Journal of Behavioral Development*, *22*, 453–473.

Ryu, E. (2008). *Evaluation of model fit in multilevel structural equation modeling: Level-specific model fit evaluation and the robustness to non-normality*. Unpublished doctoral dissertation, Arizona State University.

Ryu, E. & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, *16*(4), 583–601.

Saris, W. E., & Satorra, A. (1988). Characteristics of structural equation models which affect the power of the Likelihood Ratio Test. In W. E. Saris & I. N. Gallhofer (Eds.), *Sociometric research* (Vol. 2, pp. 222–236). London: Macmillan.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models.* Newbury Park, CA: Sage.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, *16*(4), 561–582.

Saris, W. E. & Stronkhorst, L. H. (1984). *Causal modeling in nonexperimental research: An introduction to the LISREL approach.* Amsterdam: Sociometric Research Foundation.

SAS Institute, Inc. (2002-2010). SAS Version 9.3. Cary, NC: SAS Institute, Inc.

Satorra, A., & Saris, W. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, SI,* 83–90.

Sayer, A. & Cumsille, P. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179–200). Washington DC: APA.

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: Wiley.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B, 13,* 238–241.

Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*(3), 267–288.

Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Newbury Park, CA: Sage.

Sterba, S. K. (2014). Handling missing covariates in conditional mixture models under missing at random assumptions. *Multivariate Behavioral Research*, *49*, 614–632.

Steiger, J. H. (1989). *EZPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*, 893–898.

Steiger, J. H. & Lind, J. C. (1980). Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society, Iowa City, IA.

Stoel, R. D. & van den Wittenboer, G. (2003). Time dependence of growth parameters in latent growth curve models with time invariant covariates. *Methods of Psychological Research, 8*, 21–41.

Stoel, R. D., van den Wittenboer, G., & Hox, J. J. (2004). Methodological issues in the application of the latent growth curve model. In K. van Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: Theory and applications* (pp. 241–261). The Netherlands: Kuwer Academic Publishers.

Sugawara, H. M. & MacCallum, R. C. (1993). Effect of estimation method on incremental fit indexes for covariance structure models. *Applied Psychological Measurement*, *17*, 365–377.

Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development, 37*, 240–256.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Taylor, A. B. (2008). *Two new methods of studying the performance of SEM fit indexes*. Unpublished doctoral dissertation, Arizona State University.

Tucker, L. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10.

U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS). *Public school teacher data file, 2011–12*. Retrieved from http://nces.ed.gov/surveys/sass/tables/sass1112_2013314_t1s_007.asp

Van der Leeden, R. & Busing, F. (1994). First iteration versus IGLS RIGLS estimates in two-level models: A Monte Carlo study with ML3. Unpublished manuscript, Leiden University, the Netherlands.

van de Vijver, F. J. R. & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology, 33,* 141–156.

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

Weiss, R. E. (2005). *Modeling longitudinal data*. New York: Springer.

Widaman, K. F. & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*, 16–37.

Willett, J. B. & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381.

Wu, W. (2008). *Evaluating model fit for growth curve models in SEM and MLM frameworks*. Unpublished doctoral dissertation, Arizona State University.

Wu, W. & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research, 45*(3), 420–452.

Wu,W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods, 14*(3), 183–201.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Unpublished doctoral dissertation, University of California, Los Angeles.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research, 40,* 115–148.

Yuan, K.-H. & Bentler, P. M. (2004). On chi-square difference and *z* tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement, 64,* 737–757.

Yuan, K.-H. & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology, 37,* 53–82.

Yuan, K.-H. & Hayashi, K. (2005). On Muthen’s maximum likelihood for two-level covariance structure models. *Psychometrika, 70*(1), 147–167.