

Fall 2021

Evaluating Fit Indices in a Multilevel Latent Growth Model With Unbalanced Design: A Monte Carlo Study

Fan Pan

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>



Part of the [Educational Psychology Commons](#)

Recommended Citation

Pan, F.(2021). *Evaluating Fit Indices in a Multilevel Latent Growth Model With Unbalanced Design: A Monte Carlo Study*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/6794>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

EVALUATING FIT INDICES IN A MULTILEVEL LATENT GROWTH MODEL WITH
UNBALANCED DESIGN: A MONTE CARLO STUDY

by

Fan Pan

Educational Psychology and Research
College of Education, 2021

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Educational Psychology and Research

College of Education

University of South Carolina

2021

Accepted by:

Christine DiStefano, Major Professor

Xiaofeng S. Liu, Committee Member

Jin Liu, Committee Member

Dexin Shi, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

© Copyright by Fan Pan, 2021
All Rights Reserved.

DEDICATION

I would like to dedicate this endeavor to my Dad, without whose patience this would not be possible. Thank you for your unconditional love and support, for believing in me from the very beginning, and for being there for me every step of the way. Thank you for listening to all my complaints about my entire graduate journey, for encouraging me when I needed it most, and for inspiring me to pursue my dreams. I can never express enough gratitude to you, but I hope to make you proud. This is for you!

ACKNOWLEDGEMENTS

I would first like to acknowledge my mentor and advisor, Dr. Christine DiStefano. Your guidance and continuous encouragement throughout my graduate journey mean more than I can express. Thank you for believing in me, for supporting every decision I have made, and for being there whenever I need you. Not only you encouraged me to pursue my dream and provided me with invaluable guidance, but you also helped me believe that I can finish this long way. Thank you for everything!

Also, I would like to acknowledge my dissertation committee members, Drs. Xiaofeng S. Liu, Jin Liu, and Dexin Shi. I am very thankful for all your support, for all your comments, insights, and advice, and for helping me to move forward during this process. I am really grateful for the opportunity to learn from all of you. I would also like to thank all my friends for encouraging me. Thank you!

ABSTRACT

This dissertation informed researchers about the performance of different level-specific and target-specific model fit indices in Multilevel Latent Growth Model (MLGM) using unbalanced design and different trajectories. As the use of MLGMs is a relatively new field, this study helped further the field by informing researchers interested in using various specific model fit indices.

We evaluated the descriptive information of the various model fit indices under various simulation conditions and the extent to which the fit indices could be influenced by different design factors, based on simulated data with different conditions derived from a correctly specified MLGM. Our simulation design factors included three levels of number of groups (50, 100, and 200), three levels of unbalanced GS (5/15, 10/20, and 25/75), and three trajectories (accelerating, decelerating, and linear).

Based on the results, we made recommendations for practical and theoretical research about fit indices. CFI- and TFI-related fit indices performed well in the MLGM could be trustworthy to use to evaluate model fit under similar conditions found in applied settings. However, RMSEA-related fit indices, SRMR-related fit indices, and chi-square-related fit indices varied by the factors included in this study and should be used with caution for evaluating model fit in the MLGM. The use of these fit indices appears to be particularly problematic when dealing with unbalanced design and small sample sizes.

TABLE OF CONTENTS

Dedication	iii
Acknowledgements	iv
Abstract	v
List of Tables	vii
List of Figures	viii
Chapter 1: Introduction	1
Chapter 2: Literature Review	12
Chapter 3: Method	44
Chapter 4: Results	67
Chapter 5: Discussion	91
References	109
Appendix A: Selected SAS® codes to simulate data	128
Appendix B: Sample Mplus data analysis code.....	130

LIST OF TABLES

Table 4.1 Descriptive statistics of model fit indices by NG and unbalanced GS for accelerating growth trajectory	69
Table 4.2 Descriptive statistics of model fit indices by NG and unbalanced GS for decelerating growth trajectory	75
Table 4.3 Descriptive statistics of model fit indices by NG and unbalanced GS for linear growth trajectory	79
Table 4.4 η^2 values from ANOVA design by fit index	83
Table 5.1 Recommended specific model fit indices for different model or structure of MLGM	106

LIST OF FIGURES

Figure 2.1 Linear Latent Growth Model	16
Figure 2.2 Two-level Multilevel Latent Growth Model	26
Figure 4.1 Box plot of $RMSEA_{PS_B}$ values derived from correctly specified MLGM models by NG	84
Figure 4.2 Box plot of $SRMR_B$ and $SRMR_W$ values derived from correctly specified MLGM models by NG	85
Figure 4.3 Box plot of $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, and $RMSEA_{T_S_COV}$ Values derived from correctly specified MLGM models by unbalanced GS.....	86
Figure 4.4 Box plot of $RMSEA_{PS_B}$ values derived from correctly correctly specified MLGM models by unbalanced GS.....	87
Figure 4.5 Box plot of $SRMR_W$ values derived from correctly specified MLGM models by trajectory	88
Figure 4.6 Box plot of $SRMR_{T_S_COV}$ values derived from correctly specified MLGM models by trajectory	89

CHAPTER 1

INTRODUCTION

Social science researchers are often interested in understanding how characteristics of individuals or entities change over time. These characteristics could be observations about general behavior or overall academic performance, or they could be observations about specific constructs, such as depression, communication skills, attitudes toward teachers or parents, or math ability. Longitudinal studies describe the changing pattern of characteristics of interest, allowing investigation of questions such as of how change comes about, how much change occurs, how the change process might differ across observations, and the determinants of that change over a set period. For example, using longitudinal methods, researchers interested in comparing boys' and girls' mathematics scores over primary school could gain understanding of change in mathematics scores over time and how the change may differ between genders.

Longitudinal studies have been widely applied in social sciences. In education, longitudinal methods have been used for many investigations across different content areas. For example, researchers have investigated the influence of physical education on academic achievement stability during elementary and middle school (Zhang, 2015); the relationship between depressive symptoms and school burnout during adolescence (Wen, 2012); and the early school transitions and the social behavior of children with disabilities (Wang et al., 2015).

Thus, researching methodology belongs to longitudinal study will able researchers to have a greater understanding of the changing patterns and the reasoning behind the process of change. Commonly approaches to longitudinal designs in education include evaluation of summary statistics and repeated-measures analysis of variance (ANOVA) analyses. Evaluating summary statistics provide a low-level description of the data and condense the repeated information to a single number per observation, such as mean, median, mode, and standard deviation (Muthén, 1997). Summary statistics can average response over different time points and indicate whether different subjects change in a similar or different pattern. The primary advantage of using summary statistics is that these methods are straightforward to use. The methodology and results are also easily understood by researchers who have a limited statistical background.

Repeated-measures ANOVA approaches have widely been used to analyze longitudinal data because the term “repeated” means an experimental design in which each subject is measured at several time points (Wu & West, 2010). Repeated-measures ANOVA is used to compare means of a measured attribute of different subjects and means of a measured attribute at different time points. ANOVA models test several primary null hypotheses for longitudinal research: (1) all means of measured attributes at different time points are the same; (2) all means of measured attributes in different groups of observations are the same; and (3) the groups of observations do not differ in their degree of change in the measured attributes over time. Like summary statistics, the ANOVA models have several advantages: easy to conduct, readily available in different statistical software, and the results of ANOVA models are relatively easy to understand (West et al., 2012).

Even though traditional longitudinal methods have advantages, these methods have significant limitations. For example, a significant drawback of summary statistics is that factual information is lost when all measurements are condensed into several single numbers. ANOVA models may address the primary longitudinal research question(s), but it has additional shortcomings. One limitation is that an ANOVA assumes each observation to be measured at the same time points, the methods do not analyze differences in change when observations have unequally spaced time intervals. However, when conducting empirical research, it is very hard to measure the attributes of all the observations at exactly same time points. Besides, if observations are clustered into some groups, such as students are measured over time and also clustered within classrooms, ANOVA cannot precisely account for the repeats over time. Additionally, traditional longitudinal methods do not show the change in measured attributes for each observation (Wu et al., 2008). As an example, researchers typically do not know the information of each observation at different time point with several condensed numbers from summary statistics. However, the pattern of change for individuals is typically of interest.

Modern approaches to accommodate longitudinal data have been developed to address limitations associated with the traditional methods, extending researchers' ability to describe the nature of the change of measured attributes over time (Ryu & West, 2009). Two popular classes of modern methods to analyze longitudinal data have emerged: hierarchical linear modeling (HLM) and latent growth modeling (LGM). In general, when modeling growth for longitudinal data, HLM has been regarded as a special case of LGM (Chou et al., 2004). The parameter estimates got from HLM and LGM are similar (Collins et al., 2001). HLM, which belongs to a multilevel modeling

framework, is used to analyze variance in the outcomes in nested data. Longitudinal data is considered as a kind of nested data because time points are considered as nested within each observation. HLM is very effective at explaining variance because it simultaneously investigates relationship within and between hierarchical levels of grouped data (Wu et al., 2015). LGM is another popular method. This analysis strategy belongs to the structural equation modeling (SEM) framework and considers initial status and the trajectory of growth as latent factors to model the repeated measures (Hu & Bentler, 1998). Under a LGM framework, each observation's change is represented by a unique initial status and trajectory.

Compared to traditional longitudinal approaches, HLM and LGM allow more flexibility to examine questions often posed by developmental and behavioral researchers. Traditional methods assume that the changing pattern of all individuals is the same and describe average development for the group. HLM and LGM could analyze each individual's development and show the individual variability around the overall change. HLM and LGM also analyze what factors contribute to the individual differences in individuals' growth patterns. LGM and HLM allow researchers to get around issues with assumptions when using ANOVA (Muthén & Asparouhov, 2011). For example, it is very hard to meet the assumptions about the sphericity and homogeneity of the variance structure in ANOVA. This is not a problem when using LGM or HLM. Finally, both HLM and LGM are better equipped than ANOVA at addressing unequal numbers of observations across observations.

Even though both HLM and LGM have many advantages over traditional longitudinal methods, LGMs are often preferred over HLM for several reasons.

Compared to HLM, LGM provides more flexibility for testing error structures, such as testing of homoscedasticity of measurement errors (Liu & Gould, 2002). HLM is unable to handle complex covariance structure, whereas the flexible covariance structures of LGM allows researchers to model random effects of longitudinal data. LGM allows the change to be more easily estimated, whereas HLM requires a specific form of change (Ma & Ma, 2004). So LGMs can model and comparatively evaluate a broader array of growth patterns than HLM, affording researchers greater ability to describe the nature of change (Ma & Wilkins, 2007). Further, an analytical benefit of LGM is that the researcher can only input raw data for HLM software, whereas the software for LGM can analyze covariances and indicator means.

Although LGMs provide advantages for analyzing longitudinal data, often social science data are also multilevel. Multilevel data includes different units of analysis, one nested within the other. Often, the groups are considered as one level, the observations are another level, and the measurement is the third level (Leite & Stapleton, 2011). For example, when researchers collected students' information in several schools and each student has a measurement of several time points, the students (observations) are clustered into different schools (groups), and measurements are clustered into different students. The clustering of multilevel data should be considered because observations nested within the same group have more similar characteristics than observations in different groups.

Multilevel models feature data clustering. Multilevel models test multilevel data by modeling the dependence in the data statistically, simultaneously modeling variables at different levels, without necessary recourse to aggregation or disaggregation.

Analyzing multilevel data by ignoring the nested structure will cause the standard errors of model parameters to be underestimated, leading to an overstatement of statistical significance (Langkamp et al., 2010). Standard errors for model parameters of group-level variables will be the most underestimated by ignoring the group level.

By using a multilevel model, the researchers can calculate the extent of clustering on observations' outcomes. For example, in the evaluation of school performance, research interests may concentrate on obtaining school effects on students' achievement. In a multilevel model, both the effects of group-level and subject-level variables can be estimated. In single-level approaches, the effects of group-level predictors are confounded with the effects of the subject-level predictors, and it is impossible to separate different effects due to unobserved group characteristics.

If research questions consider both change over time and data are nested, use of Multilevel Latent Growth Models (MLGM) have been advocated as a method for analysis. Combining both the benefits of multilevel models and LGM, MLGM is ideally suited for addressing the research questions concerning multilevel longitudinal data (Rappaport et al., 2016). A MLGM combines advantages of LGM (e.g., ability to incorporate indirect effects, complex measurement error structures, and multiple group analysis) while also correcting extent of clustering (Schafer, 1997). MLGM allows for both investigating observations and group trajectories within one analysis. The observation level model and group level model have different latent intercepts and latent slopes, so observation level and group level can have different growth patterns. Further, MLGMs can include characteristics of both observation and group levels to explain the

influence of various characteristics on the change patterns of two levels, the change of measured attributes of observations within each group, and the change of all observations' measured attributes (Shi et al., 2019). In addition, compared to LGM, which only considers the mean of measured attributes of time points, MLGM considers the means of different times points and the means of different groups; this can assist researchers' understanding of the overall status of the measured attributes of different groups.

As part of the broader SEM framework, MLGMs also rely upon model fit indices to determine if a tested model is a most appropriate representation of relationships in the broader population. In general, good model-data fit is as represented by fit indices within/exceeding acceptable levels, and, interpretation-wise, it is thought that the hypothesized model describes the data's structure. Poor model fit indices demonstrate the hypothesized model differs from the underlying structure observed in the dataset. Model parameters obtained from poorly fitting model cannot describe the dataset's structure and cannot, thus, cannot be used to summarize relationships between the measured attributes and observations or groups' characteristics (Simms, 2012).

When researchers are evaluating an MLGM, typical SEM model fit indices are relied upon and commonly accepted cutoff values (or “rules of thumb”) are used for interpretation (e.g., Takahashi & Wisenbaker, 1999). However, there are problems with using typical SEM model fit indices to judge the MLGM fit. In LGM, typical SEM fit indices are calculated by evaluating the entire model. Because an MLGM comprises both observation and group level models, the typical SEM model fit indices simultaneously include information from both levels. In this way, the typical SEM model fit indices are likely to be dominated by the model fit of the observation level model because the sample

size is typically much larger at the observation level than at the group model. For example, if datasets are collected from 10 schools, and each school has 200 students. The sample size for the observation level will be 10×200 , and the sample size for the group level will be 10. Large sample sizes often generate “good” traditional model fit indices. One of the popular SEM-based model fit indices Chi-square (χ^2) test is a function of sample size; thus, as sample size increases, the χ^2 will also increase. Other model fit indices are computed by using chi-square or explicitly include sample size in their calculation (Swoboda & Kim, 2010). Thus, even if the group-level model does not fit the hypothesized MLGM model, the researchers could still get acceptable model evaluation results because of the large sample size of the observation level model. As a solution to problems of typical SEM model fit indices when using MLGM, researchers have developed level-specific and target-specific model fit indices to detect whether the poor fit of the hypothesized MLGM comes from observation level model or group level model (Voight et al, 2012). Level-specific model fit indices separately evaluate different levels of MLGM and indicate whether the poor model fit comes from the observations level model or groups level model. Target-specific model fit indices tell whether the poor model fit comes from the covariance or the mean structure of observation-level model or covariance or mean structure of the group-level model. With the values from level-specific and target-specific model fit indices, researchers can fix different parts of the hypothesized model and get the MLGM with the best model evaluation.

The terms balanced and unbalanced are frequently encountered with longitudinal analysis approaches. A balanced design describes multilevel longitudinal data in which equal observations are planned to be measured at the different groups, whereas an

unbalanced design occurs when the number of observations planned to be measured at each group is not the same. It is common to encounter an unbalanced design in empirical situations. For example, states' educational policies may have a general requirement for the number of students in each class or school. The students in each class or school will fluctuate around this general number. Consider that policy states the number of students in each class to be 20; however, the actual number of students could be 18, 19, 20, or 21 per classroom.

Previous studies investigating the level-specific and target-specific model fit indices for MLGM have only examined a balanced design. Few studies concerning the MLGM model fit when unbalanced designs are present. Studies have used the effect of estimation direct maximum likelihood (direct ML) to address unbalanced issues in MLGM (Ryn & West, 2009). Direct ML conceptualizes the unbalanced design as a form of missing data. For example, if one group has 15 observations and a second group has ten subjects, the five observations can be regarded as missing data in the second group. However, direct ML can only provide traditional model fit indices for MLGM and could not output level-specific and target-specific model fit indices.

As researchers can rely on level-specific and target-specific model fit indices to judge an MLGM model's acceptability, testing if the level-specific and target-specific fit indices perform acceptably under MLGM with an unbalanced design is needed. In addition, previous studies investigating the level-specific and target-specific model fit indices with MLGM have only examined one growth pattern ((i.e., increasing growth). However, in empirical situations, different growth trajectories (e.g., constant growth, decelerating growth, and accelerating growth) may occur. As Hsu (2019) showed that

both level-specific fit indices and target-specific fit indices perform well in balanced MLGM with accelerating trajectory, these can be used to compare the model fit indices estimated from unbalanced MLGM to model fit indices from balanced MLGM with different trajectories: accelerating, decelerating, and constant.

Therefore, this dissertation study aims to fill the gaps in these literature base by informing researchers about the performance of different level-specific and target-specific model fit indices in MLGM using unbalanced design and different trajectories. There hasn't been any study of what happens under the 'best' circumstances (i.e., correctly specified model). In this study, a correctly specified MLGM was simulated considering three design factors: different group sizes, different unbalanced observation sizes, different growth trajectories parameters to investigate the performance of different model fit indices under these different conditions.

The following research questions were examined using a Multilevel Latent Growth Model with an unbalanced design.

- (1) How are level-specific and target-specific fit indices impacted by sampling error, unbalanced design, and different growth trajectories?
- (2) Do the level-specific and target-specific fit indices demonstrate reasonable sensitivity to sampling error, unbalanced design, and different trajectories?

As longitudinal approaches become more commonplace among researchers in education, the number of studies employing MLGMs has increased. MLGMs provide a flexible framework to model the changing pattern of the measured attributes in the multilevel longitudinal structure. However, even though researchers have developed the

level-specific and target model fit indices, there is also little guidance for researchers interested in using those model fit indices for unbalanced design and with different growth patterns. Considering that applied researchers may encounter differing growth trajectories, examining this condition can help to increase the study's generalizability.

CHAPTER 2

LITERATURE REVIEW

Multilevel Latent Growth Models (MLGM) are used when longitudinal data are clustered or nested data, and the objective is to investigate both individual and cluster trajectory change over time. Even though MLGMs are not yet widely applied, the model has much potential in educational research as much of the data follow a nested structure (e.g., students' data gathered from same class). Therefore, educational data are collected with cluster sampling and would benefit from MLGM.

In Structural Equation Model (SEM), model fit indices evaluate the goodness fit between a hypothesized MLGM and the collected data. However, there is currently limited research conducted about the performance of model fit indices when a MLGM is used. Typical SEM model fit indices just evaluate the whole MLGM model; however, as MLGMs have more than one level, level-specific and target-specific model fit indices have been proposed to evaluate different levels and structures of tested MLGMs. More specific fit information can help researchers understand the fit of the tested area, and, as needed to modify a MLGM. This study fills a gap in the MLGM literature by investigating the performance of different level-specific and target-specific model fit indices when evaluating unbalanced correctly specified MLGM with different sampling errors and trajectories. The results can provide guidance for researchers interested in using level-specific and target-specific model fit indices for MLGM.

This chapter presents two sections. In the first, an overview of the latent growth model is presented, defining areas such as change trajectories, unbalanced data, and model fit indices. The second section reviews studies about MLGM, involving unbalanced data analysis studies and different growth trajectories and summarizes the current practice of level-specific and target-specific fit indices with MLGMs.

THE LGM FRAMEWORK

2.1 Single-level Latent Growth Modeling

The single-level Latent growth model (LGM) framework presents the foundation for the multilevel model, as it was developed as an extension. LGMs are an advanced statistical methodology commonly used to analyze longitudinal data (Muthén & Satorra, 1995) and the model fits under the framework of structural equation modeling (SEM). LGMs examine the change pattern of an individual over time, as well as the differences in change patterns of different individuals.

A LGM design describes the growth of specific measured attributes using two latent variables: the latent intercept and the latent slope (Olejnicki & Algina, 2003). The latent intercept represents the initial level of the measured attributes at time equal to zero for a case, which is typically an individual. The latent slope describes the rate of growth for the individual, showing overall rate of change from the first measured time point to the last. With the latent slope and latent intercept, LGM examines an individual's change over time and compares change patterns across different individuals (Pornprasertmanit et al., 2013). A person's change is also called intraindividual or within-person change and the differences between persons are termed interindividual or between-person differences

(DiStefano et al, 2013). With intraindividual and interindividual change, researcher can directly compare change patterns across multiple groups and individuals.

Unlike linear regression, in which every individual is assumed to have the same intercept and slope, LGM allows these parameters to differ for individuals. LGM has other advantages over linear regression, such as the capacity to evaluate and select an appropriate model using model fit indices (Ryu, 2014). Like other SEM models, LGM has the flexibility to incorporate complex model paths and account for measurement error; LGM can also be extended to integrate exogenous covariates to account for influence on the intraindividual change or interindividual difference (Scheaffer et al., 2005). Assume that T time points of measurement, the mathematical formula of a LGM for an individual (i) is:

$$\mathbf{y}_i = \Lambda \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$

where \mathbf{y}_i is a vector, $[y_{i1}, y_{i2}, \dots, y_{iT}]$, indicating the measurements of the attribute from time 1 to time T; Λ is a $T \times 2$ matrix of factor loadings showing pattern of change over time; $\boldsymbol{\eta}_i = [\eta_{iI}, \eta_{iS}]$ is the vector of the latent intercept and the latent slope; $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \dots, \varepsilon_{iT}]$ contains the measurement errors. Measurement errors represent the degree of deviation between the observed outcome and the expected outcome for each individual.

The latent variables, $\boldsymbol{\eta}_i$, in the above formula may be expressed as functions of latent means and individual deviations away from the latent means:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i.$$

where $\boldsymbol{\alpha} = [\alpha_I, \alpha_S]$ is the mean of the latent variables; and $\boldsymbol{\zeta}_i = [\zeta_{iI}, \zeta_{iS}]$ indicates the variations in the latent intercept (I) and the latent slope (S). The latent intercept assumes the initial level of the attribute, and the latent slope in LGM describe the rate of change

over time. With the assumption of normality and homogeneity, the variation, ζ_i can be defined as:

$$\zeta_i \sim N\left(\mathbf{0}, \Psi = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix}\right), \text{ where } \varphi_{12} = \varphi_{21}$$

where φ_{11} is the variance in the intercepts, φ_{22} is the variance in the slopes, and φ_{12} is the covariance of intercept and slope factors.

Figure 2.1 defines a linear growth LGM, where change is measured at with four constant time points. IS (η_1) represents the intercept of an individual's growth trajectory and LG (η_2) represents the slope of an individual's growth trajectory. Y1-Y4 represent four measurements on the outcomes and $\varepsilon_1 - \varepsilon_4$ represent the measurement errors. Loadings on the intercept factor are fixed at 1 to represent the constant influence on the measured attributes. Loadings on the slope factor are also fixed to represent the linear increasing of the measured attributes over time; φ represents the factor variances and covariances; $\theta\varepsilon$ represents the error variances and covariances. η is the latent variable means. Under this condition, elements in the matrix Λ will be fixed and elements in the matrices φ and the matrix $\theta\varepsilon$ will be estimated.

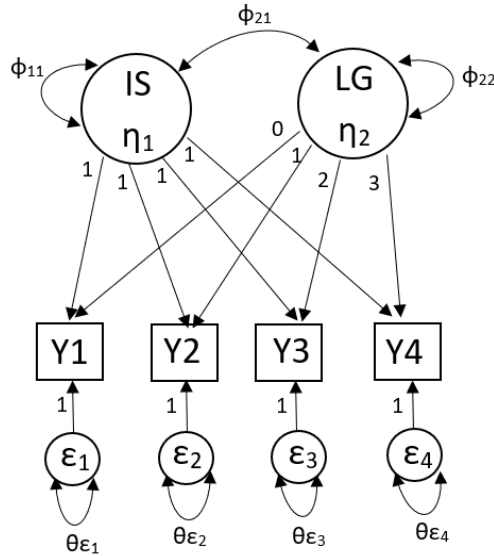


Figure 2.1 Linear Latent Growth Model

2.1.1 Measuring Change and unbalanced design

In longitudinal data analysis, LGMs estimate a trajectory that displays each individual's and overall data's responses over time (Rescorla & Rosenthal, 2004). The form of change could be linear (e.g., straight line), quadric (e.g., growth with a constant rate of rate of change), exponential (e.g., rate of change of a quantity with respect to time is proportional to the quantity itself), etc. and the direction of this change could be accelerating or decreasing. Modeling an appropriate trajectory in a LGM has important implications for accounting between-individual and within-individual difference over time (Wu & Kwok, 2012). If the 'wrong' trajectory is used, the hypothesized model will not fit the data, and the parameters estimated from the LGM do not explain the underlying structure of the data. When determining the best-fitting trajectory, researchers model the shape of trajectories from theory, where what is hypothesized is based upon

factors such as past research, theory about responses, or preliminary assessment of several individuals' changing pattern (Twisk & Vente, 2002).

In educational research, researchers commonly adopt a linear LGM, as a linear LGM is the easiest to define. This trajectory assumes a constant (i.e., linear) growth rate of all individuals to describe the change patterns across time. For example, teachers collected students' academic achievement at 6th, 7th, 8th, 9th, and 10th grades, under a linear positive growth trajectory. Each individual's outcome grows the same amount at the equally spaced intervals and have four time points as shown in Figure 1. The factor loadings in Λ would be a matrix of fixed values:

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

The numbers in the first column represent the initial score of each individual, which is the intercept. These values are all 1, indicating that each individual's intercept remains constant over the repeated measures. The second column represents the change in score per time point; numbers in this column increase from 0 to 4 with equal time intervals (no matter months/years/days), reflecting the hypothesis of linear growth. The first number in the second column was fixed at zero, thereby placing the intercept at the initial time point.

2.1.2 Model Fit Indices for LGM

Evaluating the goodness fit for a hypothesized LGM usually involves two main issues: (a) evaluating the goodness of fit between the hypothesized model and the

collected data (i.e., global fit) and (b) estimating parameters in the hypothesized model (i.e., local fit). Global fit examines the extent to which the hypothesized models, proposed based on previous theories or findings, represent the relationships among the observed variables (Kaplan, 2009; Kline, 2011). A poor fitting model indicates that the hypothesized model widely differs from the empirical data's underlying structure. Researchers typically revise a poor fitting model. Issue (b)-estimating parameters cannot be addressed in a meaningful way with a poor-fitting model. In other words, if model fit is not adequate the parameters do not summarize the relationships between the observed variables (Jöreskog & Sörbom, 1993; Woltman et al., 2012).

For a LGM, the evaluation of model fit is through typical SEM model fit indices; these indices show if the overall hypothesized LGM is suitable to summarize relationships underlying the data. Comparing the typical SEM model fit indices of different LGMs, researchers can choose the most appropriate hypothesized model. When the typical SEM model fit indices indicate a good fit of the hypothesized model, researchers usually interpret the parameters or include more variables into the model. However, if the typical SEM model fit indices show poor fit, the hypothesized model does not fit the data well, and the model is likely misspecified.

The typical SEM model fit indices belong to two types of fit indices: absolute fit indices, and relative fit indices. Absolute fit indices (e.g., chi-square test statistic (χ^2), standardized root means square residual (RMSEA) and standardized root means square residual (SRMR)) evaluate the fit of the hypothesized model without comparison to a baseline model. The typical baseline model, the null model, assumes all structural (i.e., regression) paths in the model to be zero or there is absolutely no covariance between

variables. In contrast, relative fit indices (e.g., comparative fit index (CFI) and Tucker-Lewis Index (TLI) assess the specific improvement in the fit of the hypothesized model, relative to a null model (Bentler & Chou, 1987). RMSEA, CFI, and TLI are calculated based on the χ^2 , whereas SRMR is defined through residuals (Enders, 2001a). The calculation formulas and cutoff criteria of each of the five fit indices noted above were shown in Chapter 3.

Fan and Sivo (2005) conducted a systematic review of over 200 methodological studies and noted that χ^2 , RMSEA, SRMR, CFI, and TLI are the most commonly used fit statistics for SEM. These SEM model fit indices have been proposed and widely applied to empirical LGMs studies (Enders, 2008; Enders & Bandalos, 2001). Besides, these indices have also performed well with previous simulation studies (e.g., Graham & Coffman, 2012; Gerbing & Anderson, 1992; Enders, 2006; Donoghue & Jenkins, 1992). According to Heck and Hallinger (2009), the typical SEM model fit indices are widely used in LGM because they are available in most statistical programs.

When testing an LGM, as the typical SEM model fit indices evaluate the global or whole model simultaneously and do not detect a misspecified structure. A misspecified structure can have biased coefficients or error terms and tend to have biased parameters estimates. In this way, the results from typical SEM model fit indices do not provide researchers information about how to modify a hypothesized LGM model.

Target-specific Model Fit Indices for LGM

Schermelleh-Engel et al. (2014) proposed that target-specific fit indices can detect a misspecified structure of LGM, including misspecifications in a mean or covariance structure. Target-specific fit indices were suggested to evaluate these two structures

separately. Misspecifications in mean structure of LGM can be misspecified marginal mean structure and misspecified conditional mean structure. The misspecified marginal mean structure indicates the misspecification of mean growth trajectory, and misspecified conditional mean structure refers to the misspecification of individual growth trajectories (Jamshidian & Mata, 2008). Misspecifications in covariance structure of LGM can be a misspecified within-individual covariance structure and misspecified between-individual covariance structure (Wu et al. 2009). The misspecified between-individual covariance structure means the misspecification of variances and covariances among growth trajectories. The misspecified within-individual covariance structure means a misspecification of variances and covariances of residuals, which is part of the data that cannot be explained by the hypothesized LGM model (Wu et al., 2009).

Wu and West (2010) first calculated different target-specific fit indices for a LGM. Target-specific fit indices for the mean structure (T_S_Mean) of LGM include $CFI_{T_S_Mean}$, $TLI_{T_S_Mean}$, $RMSEA_{T_S_Mean}$, and $SRMR_{T_S_Mean}$, and target-specific fit indices for the covariance structure (T_S_COV) of LGM include $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, $RMSEA_{T_S_COV}$, and $SRMR_{T_S_COV}$. T_S_Mean fit indices are calculated by saturating the covariance structure, whereas T_S_COV fit indices can be calculated by saturating the mean structure. A saturated structure is a just identified structure, meaning that the number of proposed parameters exactly equals the number of known values of the data. In a LGM with saturated covariance structure, the covariance between latent intercept and latent slope is defined as 0, and the means of latent intercept and latent slope are freely estimated. As for a LGM with saturated mean structure, the covariance between latent intercept and latent slope is freely estimated, and the means of latent intercept and latent

slope are defined as 0. In this way, a saturated mean or covariance structure in LGM has zero degrees of freedom, and the chi-square test statistics of a saturated structure equals zero (Schumacker & Lomax, 2010).

Several studies have examined the sensitivity of target-specific fit indices for LGM. Wu and West (2010) compared the target-specific fit indices with traditional fit indices. Investigating the overall fit of the LGM model, they indicated that target-specific fit indices show which structure (mean or covariance) contributed to the overall misspecified model. The LGM simulated by Wu and West (2010) was an accelerating growth model with five time points. This study also simulated a misspecified latent mean structure, a misspecified latent variance structure, and a combination of both. Four sample sizes were considered: 125, 250, 500, and 1000. Compared to the traditional fit indices, T_S_Mean fit indices showed more sensitivity to misspecifications in the mean structure, because saturating the covariance structure dramatically reduced freedom degrees. The lower degree of freedom available for LGM increases the power of T_S_Mean fit indices (Wu & West, 2010). However, as for detecting the misspecifications in the covariance structure, T_S_COV fit indices did not show substantial power and performed similarly as compared to the traditional fit indices. This is because saturating the mean structure only decreases the degrees of freedom by a small amount (Wu & West, 2010). This study indicated that T_S_Mean fit indices can identify the misspecified mean structure of LGM, whereas T_S_COV fit indices cannot identify the misspecified covariance structure of LGM.

Raykov and Marcoulides (2012) investigated three different growth rates and five different sample sizes with a single-level LGM. The growth rate included a linear, accelerating, and piece wise trajectory with five time points and sample sizes included were: 100, 200, 500, 1000, and 2000. The goal of the study was to compare the performance of target-specific fit indices with typical SEM fit indices applied to a LGM. Roberts and Bryant (2011) showed similar results with Wu and West's findings, confirming that $RMSEA_{T_S_COV}$ did not improve the power of detection of misspecified covariance structure when compared to the typical RMSEA index. Findings for the $SRMR_{T_S_MEAN}$ suggested higher power for rejecting misspecifications in the mean structure than typical SRMR. Both typical Chi-square (χ^2) and $\chi^2_{T_S_COV}$ performed well to identify a misspecified growth shape. The results also indicated that $RMSEA_{T_S_COV}$ has higher sensitivity over other model fit indices to detect a misspecified model when the growth rate is nonlinear. These two studies suggested that T_S_Mean fit indices performs better than typical SEM model fit indices, whereas T_S_COV fit indices performs similar with typical SEM model fit indices.

INTRODUCING THE MLGM FRAMEWORK

2.2 Multilevel Latent Growth Modeling

Studies in education are typically multilevel because the sampling units are typically nested, such as selection of schools or classrooms instead of individual students. For example, most large-scale longitudinal studies commonly adopt two-stage cluster sampling (TCS) (e.g., the Education Longitudinal Study of 2006; or the Early Childhood Longitudinal Study, Shin et al., 2009; Kindergarten Class of 2004–2005, Shin et al.,

2013). The first stage of TCS includes a random selection of clusters, and the second stage consists of selecting subjects from each cluster. Often, researchers first select several schools and collect students' information in these schools (Savalei & Bentler, 2009).

The multilevel latent growth model (MLGM) framework has been widely used to study longitudinal multilevel data (Muthén, 1994; Yuan & Bentler, 2000; Schermelleh-Engel et al., 2003). By randomly selecting clusters (e.g., schools or classes) and then randomly selecting subjects (e.g., students) within the selected clusters (Holder, 2015), TCS makes the research design more efficient (Morgan et al, 2011). However, because of the TCS, subjects in the same cluster usually have some degree of dependence among observations (Poon & Lee, 1992). For example, students in the same classroom tend to be more alike due to the effects of the same learning environment.

Ignoring the dependent nature of the data and carrying out an LGM analysis for multilevel longitudinal data could lead to severe distortions of model fit and standard errors of estimates (Muthén, 1997). LGM uses the conventional maximum-likelihood covariance structure analysis, which assumes simple random sampling, to estimate the parameters and calculate model fit (Kim, 1990). However, if dependence between the observations exists, acceptable model fit values may be outputted, even though the hypothesized model does not fit the data. The standard errors of parameter estimates are also attenuated as model fit values when there is dependence. Without including the hierarchical nature of multilevel data, researchers may incorrectly retain a LGM model.

MLGM extends the LGM model by accommodating the dependence between observations due to nested data. The nested longitudinal data include repeated measures

for each individual nested within the groups, thus forming a three-level structure. Based on previous research (e.g., Judge & Watson, 2011; Kromrey & Hines, 1994), the three-level structure can be specified with a two-level model. In this two-level model, subject-related parameters are estimated in the within-level model, and cluster-related parameters are evaluated in the between-level model. MLGM can output different parameters for different levels in the model, allowing researchers to separately study different levels. MLGM also permits a comprehensive investigation of different trajectories for the subjects and clusters (Muthén & Asparouhov, 2011). Factor loadings at different levels are set to be equal to obtain unbiased parameter estimates and statistical inferences (Muthén, 1997). Specific factor loadings for MLGM can relax to accommodate nonlinear growth and allow the cluster's and subject's growth rates to vary across time points. Like LGM, MLGM can also adjust exogenous covariates to account for a covariate's influence on both the trajectories of between-level and within-level. With the trajectories for clusters and subjects, researchers can design different interventions at the cluster and/or subject level.

Considering a multilevel longitudinal data with T time points of measurement for each N subjects clustered within G groups. The between-level effects of the g^{th} group can be modeled as:

$$\mathbf{y}_g = \Lambda \mathbf{B} \boldsymbol{\eta}_g + \boldsymbol{\varepsilon}_g ,$$

In the equation, $\boldsymbol{\mu}_g = [\mu_{g1}, \dots, \mu_{gT}]$ describes the intercepts of longitudinal measurements for the g^{th} cluster; Λ_B is a $T \times 2$ matrix including the cluster-level factor loadings; $\boldsymbol{\eta}_g = [\eta_{gI}, \eta_{gS}]$ contains the growth parameters for latent intercept and slope variables for the g^{th} cluster; $\boldsymbol{\varepsilon}_g = [\varepsilon_{g1}, \dots, \varepsilon_{gT}]$ is the cluster-level error; $\boldsymbol{\alpha} = [\alpha_I, \alpha_S]$ describes the mean of

latent growth parameters; and Σ_B is a $T \times T$ matrix describing the variation in group differences.

For the i^{th} individual within the g^{th} group, the within-level responses can then be modeled as:

$$\mathbf{y}_{gi} | \boldsymbol{\mu}_g = \boldsymbol{\mu}_g + \Lambda_w \boldsymbol{\eta}_{gi} + \boldsymbol{\varepsilon}_{gi}$$

Here, $\mathbf{y}_{gi} = [y_{gi1}, \dots, y_{giT}]$ describes the measurements of the i^{th} individual in the g^{th} cluster; Λ_w represents the $T \times 2$ subject-level factor loading matrix; $\boldsymbol{\varepsilon}_{gi} = [\varepsilon_{gi1}, \dots, \varepsilon_{giT}]$ means the subject-level error; $\boldsymbol{\eta}_{gi} = [\eta_{gi1}, \eta_{gi2}]$ denotes the latent growth parameters for the i^{th} individual nested in the g^{th} cluster; and Σ_w is a $T \times T$ matrix containing the variation in individual differences.

Figure 2.2 shows a two-level linear growth MLGM with four constant growth time points. IW (η_1) represents the intercept of an individual's growth trajectory and LW (η_2) represents the slope of an individual's growth trajectory. Y1-Y4 represent four continuous outcomes for individuals, and $\varepsilon_1 - \varepsilon_4$ represent the degree of deviation between the observed outcome and the expected outcome of individuals. Λ represents the factor loading for individual-level; ϕ represents the factor variances and covariances for individual-level; $\theta\varepsilon$ represents the error variances and covariances for individual-level. η is the latent variable means for individual-level.

IB (η_1) represents the intercept of a group's growth trajectory and LB (η_2) represents the slope of a group's growth trajectory. Y1-Y4 represent four continuous outcomes for groups, and $\varepsilon_1 - \varepsilon_4$ represent the degree of deviation between the observed outcome and the expected outcome for groups. Λ represents the factor loading for group-level; ϕ represents the factor variances and covariances for group-level; $\theta\varepsilon$ represents the

error variances and covariances for group-level. η is the latent variable means for group-level. Under this condition, both group-level and individual-level's matrices Λ will be fixed. The matrices ϕ and the matrix $\theta\epsilon$ of both levels will be estimated. Typically, factor loadings of different levels are set to be equal to obtain unbiased parameter estimates and statistical inferences (Muthen, 1997).

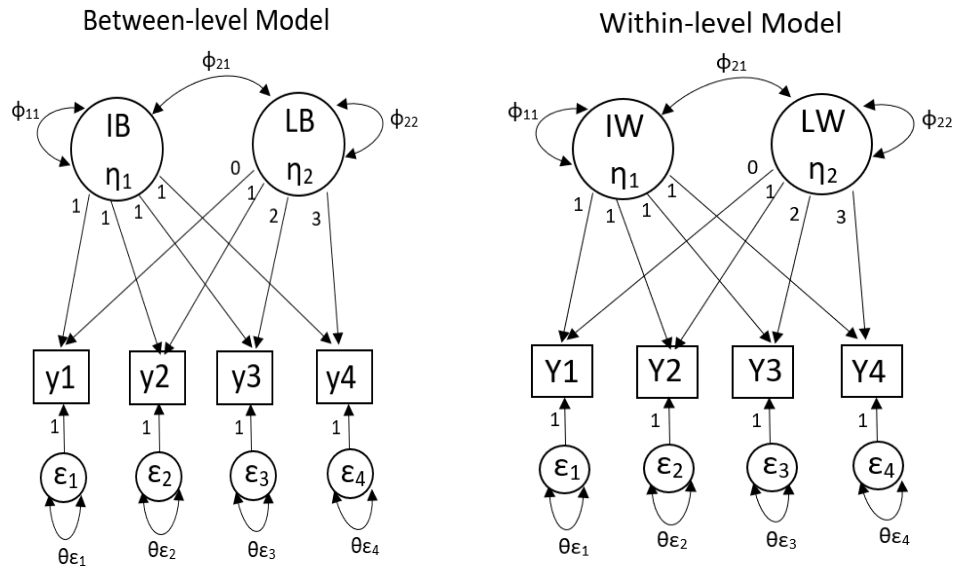


Figure 2.2 Two-level Multilevel Latent Growth Model

2.2.1 Investigating Fit of MLGM Designs.

Muthén (1997) compared the model estimation results of MLGM, such as model fit indices and standard errors of parameters, to results of LGM, single-level conventional autoregressive model, and two-level conventional autoregressive model. The study investigated empirical multilevel, longitudinal data from Longitudinal Study of American Youth, a national longitudinal data of students' mathematics and science education in U.S. public schools (LSAY; Mundform et al, 2011). The MLGM proposed in Muthén's

(1997) study was a two-level model with four time points. Compared to other models, such as LGM and multilevel Confirmatory Factor Analysis, MLGM computed the best model evaluation information, indicating that MLGM was the appropriate hypothesized model for multilevel longitudinal data.

Even though MLGM could extend benefits of the LGM model by accommodating the correlation between observations, applications of MLGM in longitudinal research are not widespread in many areas; however, examples of this method do exist in education. For example, Muthén (2004) conducted a two-level MLGM to study school effects among different social class composition schools. The data consisted of 779 public high schools drawn from National Education Longitudinal Study of 1988 and 5120 students and divided into three group: low, middle, and high social class (SES/poverty). The achievement values of all students in 779 schools were collected at 8th, 10th, and 12th grades. Results showed the effects of school characteristics and practices on change of schools' achievement values (i.e., mean of all students' achievement values in each school) differed across the three social class types.

MLGMs were used to examine the effects of school characteristics and practices on change of students' achievement values. Schweig (2014) used a two-level MLGM to investigate the impact of African American paternal figures, including paternal presence and warmth, on their daughters' psychological adjustment and educational outcomes. This study included over 2000 female students from ages 11-16 from an extremely impoverished community; data were collected across 10 years of observation from 2000 to 2009. Two outcomes were measured: girls 'psychological adjustment and girls' educational outcomes. The MLGM results showed that paternal figures did not

significantly influence girls' psychological adjustment but significantly impact girls' educational outcomes. More recently, Sessoms (2019) conducted a two-level MLGM with three time points to investigate the change of the perceived stress caused by significant psychological task change over time. The longitudinal data in the research was a sample of 393 beginning teachers. With different levels of MLGM, Sessoms (2019) study depicted the growth of stress cause and stress response. In the within-level of MLGM, individual growth trajectories were estimated to show the change of perceived stress over time points for each individual. Between-level measured the differences in change over time across individuals, determining the effects of perceived stress on difference between trajectories of individuals. The results indicated that both the stress causes and stress responses changed over time, but that specific intervention reduced the change of stress causes and stress responses.

2.2.2 Model Fit Indices for MLGM

Applied SEM researchers have relied heavily on typical SEM fit indices and the cutoff values to determine model-data goodness fit and model results. (e.g., Shin et al., 2013; Voight et al, 2011; Sirin, 2005; Swoboda & Kim, 2010). However, except for the overall model chi-square, other SEM fit indices do not appropriately evaluate the fit of an MLGM because the multilevel structure is not considered. Researchers need a model evaluation method that can separately estimate each level and different structure of MLGM to understand which part of the hypothesized model did not fit the data.

As an MLGM contains both between-level and within-level models, the typical SEM fit indices to assess the model fit have a potential limitation in detecting the between-level model's lack of fit. Because the sample size is much larger at the within-

level than at the between-level, the typical SEM fit indices are likely to be dominated by the level with a larger sample size (Roberts & Bryant, 2007; Schumacker & Lomax, 2010; Raudenbush & Bryk, 2002). In this way, the typical SEM fit indices are more sensitive to misspecification in the within-level model (Hsu et al, 2015). For example, if typical SEM fit indices indicate an excellent fitting model, a researcher does not know if both the between-level and within-level model fit well or because the typical SEM fit indices fail to detect the misspecification at the between-level (Kwon, 2011). In addition, as typical SEM fit indices output if the overall fit of a hypothesized model to a set of data, researcher does not know if the misspecification is due to the covariance matrices or mean structure. In this way, the results from both typical SEM fit indices do not provide researchers information about how to modify the hypothesized model.

When using MLGM, the level-specific fit indices and the target-specific fit indices have also been proposed to overcome the above limitations of the typical SEM fit indices. Level-specific fit indices for MLGM individually estimate each level model and are very informative in locating the source of model misspecification for MLGM: between-level model or within-level model (Kwon, 2011). Target-specific fit indices for MLGM examine whether the whole model's misspecification comes from the covariance structure or the mean structure of between-level or within-level model. Similar to target-specific fit indices for LGM, target-specific fit indices for MLGM could also investigate the sensitivity of the different model misspecifications, such as omitting a factor loading or a factor covariance from the model within the SEM (Lüdtke et al, 2010).

Level-specific fit indices for MLGM

A partially saturated model (PS model) has been proposed to obtain the level-specific fit indices (Marsh et al, 2005). A PS model means that in an MLGM, either a within-level model or a between-level model is a saturated model. Different from a saturated model, which defines whole model is a just identified model, the PS methods defines the between-level model as a just identified model and keeps the within-level model as hypothesized model or defines the within-level model as a just identified model and keeps the between-level model as a hypothesized model. A PS model can be obtained by correlating all the observed variables and allowing all the covariances or correlations to be freely estimated at the between-level or within-level model. In saturated within-level model, the covariance between IW (η_1) and LW (η_2) in MLGM formula is freely estimated, and the covariances between four outcomes for individuals, Y1-Y4, are also freely estimated. As for saturated between-level model, the covariance between IB (η_1) and LB (η_2) in MLGM formula is freely estimated, and the covariances between four outcomes for groups, Y1-Y4, are also freely estimated. Neuman (2009) demonstrated that the PS method provides level-specific fit indices with reasonable non-convergence rates and Type I error rates. Non-convergence rates examine the number of replications failed when obtaining 1000 converged solutions and was the first step in the analysis of the results. With a high non-convergence rate, a model fails to achieve equilibrium during analysis (Nakai, 2011). Ryu and West (2009) indicated that PS model generated low non-convergence rate and was appropriate used to generate level-specific fit indices.

Using the PS method, Ryu and West calculated the between-level specific fit indices (PS_B) and within-level specific fit indices (PS_W), meaning that different levels

will be evaluated by different fit indices (Hsu et al, 2015; Ryu & West, 2009). The between-level specific χ^2 ($\chi^2_{PS_B}$) can be calculated by specifying a hypothesized between-level model and saturating the within-level model because the saturated within-level model's χ^2 equals zero and do not contribute to the χ^2 of the whole model (Hsu et al, 2016). In this way, the $\chi^2_{PS_B}$ only reflects the model fit of the hypothesized between-level model (Hsu et al, 2016). After $\chi^2_{PS_B}$ is obtained, other between-level specific fit indices, such as $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} , can be calculated because these fit indices are based on the χ^2 . Similarly, within-level specific χ^2 ($\chi^2_{PS_W}$) can also be obtained by specifying a hypothesized within-level model and saturating the between-level model (Grimm, 2008). After $\chi^2_{PS_W}$ is obtained, other within-level specific fit indices, such as $RMSEA_{PS_W}$, CFI_{PS_W} , and TLI_{PS_W} , can also be computed. Statistical packages for use with MLGM, such as Mplus, can calculate these fit indices.

Previous literature, investigating level-specific fit indices' performance, has been conducted for different multilevel SEMs (MSEM): multilevel confirmatory factor analysis, multilevel path models, multilevel nonlinear models, and MLGM. (Hsu et al., 2016; Enders, 2001b; Schermelleh-Engel et al., 2014). Ryu and West (2009) simulated a multilevel confirmatory factor model to investigate the level-specific evaluation. The within-level had six continuous observed indicators. Three of the six indicators loaded on one within-level factor, and the other three indicators loaded on another factor. The between-level had the same measurement structure as the within-level. This research considered the balanced design and different ratios of within-level and between-level samples. The sample sizes simulated for the between level varied from 30, 50, to 100 when within level sample size was held to 100. Ryu and West (2009) also simulated

sample sizes for within-level as 50, 200, and 1000, while the between-level samples were controlled as 50. Ryu and West (2009) indicated that within-level specific fit indices correctly indicated the within-level model's poor model fit, and between-level specific fit indices successfully detect the lack of fit in the between-group model.

Based on Ryu and West (2009) results, Ryu (2014) illustrated the level-specific model evaluation using empirical data and provided recommendations to researchers interested in using MSEM. A two-level path model was hypothesized for the collected data, and two other misspecified models based on the hypothesized model were also estimated. In one misspecified model, the relationship between two variables in the between-level was fixed as zero. The relationship between two other variables in the within-level was also fixed as zero for the other misspecified model. The level-specific fit indices indicated poor fit for the two misspecified models. This empirical study substantiated the results of Ryu and West's (2009) simulation study, showing that level-specific model evaluation provides researchers valuable information by assessing the model fit separately as each level.

As an extension of the above two studies, Hsu et al. (2016) considered the impact of intraclass correlation coefficients (ICCs) on the performance of level-specific fit indices in simulated MSEM. The ICC is defined as the ratio between group-level variance and total variance (Peugh & Enders, 2004). The ICC identified as larger than 0.15 are common in educational research and showed that the clustering should be considered (Hox, 2010). Following Ryu and West's (2009) study, Hsu et al. (2016) also simulated a two-level MSEM with between-level and within-level factors with six continuous observed indicators loaded on these factors. Savalei and Bentler's (2005)

findings showed that when the within-level sample size was 50, non-convergence problems occurred. To overcome this limitation, Hsu et al. (2016) changed the condition of the within-level sample size of Ryu and West's (2009) study to 100, 200, and 500; however, the focus here was on the within-level given that as Schafer and Graham (2002) discovered that the between-level sample size was not an influential factor for MSEM evaluation, Hsu et al. (2016) used 30, 50, and 100 as between-level sample size just as Ryu and West's (2009) study. Hsu et al.'s (2016) results showed that the ICC does not significantly affect the effectiveness of level-specific model fit indices. When ICC was very low, CFI_{PS_W} and TLI_{PS_W} can still detect the misspecification for between-level models, whereas $SRMR_B$ and $RMSEA_{PS_W}$ did not work. ICC did not influence all within-level fit indices.

Only one study to date has concentrated on the level-specific fit indices in MLGM (Hsu, 2019). In line with Wu and West's (2009) study, this simulation study extended Wu and West's (2009) single level LGM to a two-level MLGM model with the same accelerating quadratic trajectory and time points. The estimated MLGM had five time points, and each time point was assumed to be on a standardized scale (i.e., $M = 0$ and $SD = 1$). The parameter settings were simulated based on empirical data from the Longitudinal Surveys of Australian Youth (LSAY); This dataset collects information about education, training, financial matters, health, social activities, and related issues from a large sample of students. Following Wu, Kwok, and Willison's (2015) simulation study, the number of clusters (NC) for Hsu (2019)'s study was simulated as 50, 100, 200 and cluster sizes (CS) were designed into three levels, 5, 10, and 20. The results showed that CFI- and SRMR- related fit indices were not affected by small NC or CS. The

RMSEA- related fit indices were likely to be influenced by small NC or CS. TLI- related fit indices needed a moderate NC (100) and CS (10). The results also indicated that between-level specific fit indices, $RMSEA_{PS_W}$, CFI_{PS_W} , and TLI_{PS_W} , were not sensitive to the misspecified between-covariance structure, whereas $SRMR_B$ was recommended to detect this misspecification. As for the misspecified between-mean structure, $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} were suggested. Among them, $RMSEA_{PS_B}$ was recommended as it was found to be more sensitive to detecting misspecification.

Previous studies investigating MLGM have only simulated balanced datasets and have examined only one growth trajectory. Even though level-specific fit indices performed well in these studies, unbalanced datasets and other growth trajectories are needed to be modeled.

Target specific fit indices for MLGM

In addition to level-specific fit indices, our research also evaluated the performance of target-specific fit indices. Level-specific fit indices for MLGM individually estimate each level model. Target-specific fit indices for MLGM examine whether the misspecification comes from the covariance structure or the mean structure of between-level or within-level model. Although target-specific fit indices were recommended to determine the misspecified structure of LGM, the literature has not yet provided adequate empirical evidence to support the use of target-specific fit indices for MLGM. As noted earlier, there is only one mean or covariance structure in LGM, so the misspecification can only occur in either mean or covariance structure. However, in MLGM, both the between-level and the within-level model have a covariance structure and a mean structure. Hsu et al. (2019) extended the investigation of target-specific fit

indices' performance from the context of LGM to MLGM. The authors outlined a practical way to compute the target-specific fit indices for the between-level covariance structure fit indices and the between-level mean structure fit indices. The target-fit indices for MLGM only need to be estimated at the between-level model. Because the means of growth factors are fixed at zero, the misspecifications of the whole MLGM could only be attributed to the within-covariance structure (Muthén, 1997). The fit indices for between-level covariance structure include $\chi^2_{T_S_COV}$, $RMSEA_{T_S_COV}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, and $SRMR_{T_S_COV}$, and the fit indices for the between-level mean structure has $\chi^2_{T_S_Mean}$, $RMSEA_{T_S_Mean}$, $CFI_{T_S_Mean}$, $TLI_{T_S_Mean}$, and $SRMR_{T_S_Mean}$.

Based on Wu and West's (2010) and Ryu and West's (2009) research, Hsu et al. (2019) generated T_S_MEAN fit by saturating the within-level model and the covariance structure of the between-level model. T_S_COV fit indices were created by saturating the within-level model and the mean structure of the between-level model. The researchers studied the influence of the sample size, cluster size, and type of misspecification on the sensitivity of target-specific fit indices for MLGM. The results indicated that $RMSEA_{T_S_COV}$, $CFI_{T_S_COV}$, and $TLI_{T_S_COV}$ showed higher sensitivity to misspecified between-variance structure than $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} . In addition, the $RMSEA_{T_S_COV}$ yielded a higher sensitivity than the other two fit indices. $\chi^2_{T_S_COV}$ is also favored because of its high power in different sample size conditions. $SRMR_{T_S_COV}$ is not recommended when the cluster size is less than 5. As for a misspecified between-mean structure, $RMSEA_{T_S_Mean}$, $CFI_{T_S_Mean}$ and $TLI_{T_S_Mean}$ did not show a higher sensitivity than $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} . Hsu et al. (2019) recommended researchers use $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} to detect misspecified between

structures. Both $\text{SRMR}_{T_S_mean}$ and SRMR are not recommended because they had means and variances close to 0.

Previous studies investigating target-specific fit indices indicated that those fit indices performed well in MLGM have with balanced datasets and accelerating growth trajectory. Unbalanced datasets and other growth trajectories are needed to be modeled to provide researchers a detailed guide about using target-specific fit indices.

2.2.3 Trajectory for MLGM

Like LGM, MLGM employs trajectories to measure the change of observed attributes in different levels and use factor loadings to define trajectories. However, factor loadings of latent intercept and slope at different levels are set to be equal to obtain unbiased parameter estimates and statistical inferences (Hu & Bentler, 1999). Trajectories can be modeled as linear or nonlinear. For example, in a two-level MLGM with accelerating quadratic growth at both levels, the random vectors of latent growth factors η of both levels contains the growth parameters for the latent intercept, linear slope, and quadratic slope variables. The factor loadings for between-level Λ_B and within-level Λ_W for T (e.g., five) time points are set as same:

$$\Lambda_B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \quad \Lambda_W = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}$$

To determine conditions typically faced by empirical researchers, over 1000 peer-reviewed articles education were reviewed. The articles were searched with the key words “multilevel” and “education” in all databases in University of South Carolina’s library. All these studies model the data with multilevel nature (e.g., students are sampled

by district, school, or class), even though some of these studies did not consider the multilevel data structure in the analysis. Among those 1000 studies, 30 studies conducted MLGM to investigate empirical data. The areas studied from these MLGM studies include children's general academic ability, bilingual education, reading and math education, and parental involvement, etc. In the analyses, researchers reported using constant, decelerating, and accelerating trajectories. Among the 30 studies, 14 modeled the growth trajectory as linear. Among the 16 studies that reported a nonlinear constant growth rate, ten found that the growth trajectory of the model decelerating, four studies assumed the growth trajectory to be accelerating, and two researchers found that the growth trajectory to be linear. Ten studies added a quadratic growth rate to the linear growth trajectory. Four studies considered the trajectory as unspecified, meaning that the growth rate was freely estimated. Four studies fit a piecewise linear growth model, allowing the model could have two different growth rates for two separate periods. There was one study, which specifies the growth trajectory to be a natural log of time (i.e., relative growth rates are $\ln(1)$, $\ln(2)$, ..., $\ln(t)$; Shin et al., 2013).

Previous empirical studies have simulated MLGM with different trajectory growth patterns to investigate the effects on outcomes. For example, Lee et al (2010) simulated an accelerating quadratic trajectory and fit a misspecified MLGM model to investigate the model fit indices' sensitivity. Hsu (2019) simulated a correctly specified and five different misspecified MLGM with accelerating quadratic trajectory to determine the effectiveness of level-specific and target-specific fit indices. As for decelerating quadratic trajectory, Lenkeit (2012) simulated multilevel mixture models with decelerating quadratic trajectory to evaluate the usefulness of model fit indices for

estimating multilevel longitudinal data. Wu and West (2010) indicated that between-level specific fit indices, $RMSEA_{PS_W}$, CFI_{PS_W} , and TLI_{PS_W} , were not sensitive to the misspecified between-covariance structure in an MLGM model with an accelerating quadratic trajectory. As for the misspecified between-mean structure, $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} were suggested in an MLGM model. The Hsu (2019)'s results showed that, in a simulated MLGM with an accelerating quadratic trajectory CFI- and SRMR- related fit indices were not affected by sample size, the RMSEA- related fit indices were likely to be influenced by sample size, and TLI- related fit indices needed a moderate sample size. Lenkeit (2012) indicated that both within-level specific fit indices and global model fit indices can correctly indicated the within-level model's poor model fit indices under. However, between-level specific fit indices performs better than global model fit indices when detect the lack of fit in the between-group model of multilevel mixture models with decelerating quadratic trajectory.

All these above studies tried to figure out the effectiveness of the model fit indices under commonly used MLGM model with one type of trajectory. However, even though these previous studies showed that some model fit indices worked well for MLGM with one specified trajectory, there is no studying comparing the performance of model fit indices across different trajectories. This study have filled this gap, trying to figure out if trajectory influence the effectiveness of different model fit indices.

2.2.4 Unbalanced Design for MLGM

As with other longitudinal studies, researchers who adopt MLGMs will most likely encounter an unbalanced design, as this design can be a common phenomenon over multiple waves of data collection (Li, 2010). An unbalanced MLGM design is different

from the definition of an unbalanced LGM design. For LGM, the unbalanced is defined as individuals measuring with unequal intervals. In multilevel longitudinal situations, unbalanced data typically refers to situations where the number of subjects per cluster is not equally distributed (May & Supovitz, 2006). That is, some clusters could have more subjects, whereas other clusters could have fewer subjects.

For a balanced MLGM with G balanced groups, each group has n observations. The total sample size N equals nG . The MLGM defines the within group covariance matrix as S_{PW} and the between group covariance matrix as S^*_B . The formulas for the S_{PW} & S^*_B covariance matrices are:

$$S_{PW} = \frac{\sum_g \sum_i^n (Y_{gi} - \bar{Y}_g)(Y_{gi} - \bar{Y}_g)'}{N - G}$$

$$S^*_B = \frac{\sum_g^n n(\bar{Y} - \bar{Y}_g)(\bar{Y} - \bar{Y}_g)'}{G - 1}.$$

In the above two equations, Y_{gi} represents for the response for each observation, \bar{Y}_g represents the mean response of n observations in each group, and \bar{Y} indicates for the mean response of all N observations in the data.

In an unbalanced MLGM situation, as groups have unequal numbers of individuals, S_{PW} may still represent the within group covariance matrix because the S_{PW} formula directly pools together all observations, regardless of group size. S^*_B , however, cannot represent the covariance matrix for each group because each group could have a distinct group size, n . Different S^*_B matrices will be calculated for each group. In this way, the aggregate covariance for unbalanced multilevel data no longer represents sufficient statistics for model estimation and may cause problems for model estimation.

Unbalanced multilevel longitudinal data may be prevalent in education for many reasons. For example, in the National Education Longitudinal Study (NCES, 1993), different elementary schools have different numbers of students because of different funding source, community environment, etc. The number of students per elementary school at different time points is also different, because students can drop out of school or transfer from or to other schools.

Although unbalanced multilevel longitudinal data is common with many educational research applications, there is not yet a study that has investigated the influence of unbalanced multilevel longitudinal data on model fit indices of MLGM. Jackson (2003) simulated an unbalanced multilevel data at one time point to investigate the performance of a multilevel confirmatory factor model on unbalanced data. The results indicated that unbalanced data had little impact on the accuracy of parameter estimates of the within (i.e., individual) level model. However, for the between (group) level, the variances of model fit indices tended to be underestimated, and standard errors of parameter estimates were too small. As the Graham (2003) investigation used one time point, results are not thought to improve with a MLGM situation. Model estimation for MLGM, including model fit, parameters estimation, and standard errors, may be substantially biased if the unbalanced nature is not considered.

This study examined the performance of different level-specific and target-specific model fit indices when evaluating unbalanced MLGM with different sampling errors and trajectories. With balanced data, as each cluster has constantly measured attributes, one covariance and one mean structure could represent the relationship between subjects within each cluster. The mean structure of balanced data is calculated

by summing the values of all individuals in the cluster to divide the fixed number of individuals. However, as each cluster in unbalanced data do not have same numbers of subjects, one mean structure could not stand for mean structure of all clusters. Each cluster's covariance is also different due to different numbers of measured attributes. The non-constant covariance structure within each level may cause a severe concern, especially if separate trajectories for subjects and clusters are of interest (Harmsen et al., 2019). Different number of subjects in each cluster may cause misspecification of mean and covariance structures for each level required by the model estimation (e.g., Graham et al 1996; Gibson & Olejnik, 2003) and result in low statistical power for overall MLGM estimation (Goos et al., 2013).

Previous studies investigating both level-specific and target-specific fit indices for MLGM have only simulated balanced datasets. There is no research model MLGM with unbalanced design. As unbalanced design is very common in educational research, unbalanced datasets are needed to be studied to provide researchers a detailed method about using these fit indices.

In summary, as an MLGM contains both between-level and within-level models, the typical SEM fit indices have limitation in detecting the between-level model's lack of fit. Previous studies proposed the level-specific fit indices and the target-specific fit indices for MLGM to overcome the limitations of the typical SEM fit indices. Level-specific fit indices locate the source of model misspecification for between-level model or within-level model. Target-specific fit indices examine whether the model's misspecification comes from the covariance structure or the mean structure of between-level. The target-fit indices for MLGM only need to be estimated at the between-level

model. Because when the means of growth factors are fixed at zero, the misspecifications detected by the within-level specific model fit indices will be due to the covariance structure of within-level model. A partially saturated model has been proposed to obtain the level-specific fit indices and target-specific fit indices. The between-level specific model fit indices can be calculated by specifying a hypothesized between-level model and saturating the within-level model. Similarly, within-level specific model fit indices can be obtained by specifying a hypothesized within-level model and saturating the between-level model. Target-specific fit indices for mean structure can be generated by saturating the within-level model and the covariance structure of the between-level model. Target-specific fit indices for covariance structure fit indices can be created by saturating the within-level model and the mean structure of the between-level model. Based on previous studies, we conclude that when evaluating a balanced MLGM with and using an accelerating trajectory, CFI_{PS_B} and $SRMR_{PS_B}$ worked best for evaluating between-level model, and CFI_{PS_W} and $SRMR_{PS_W}$ work best for evaluating within-level model. $RMSEA_{T_S_COV}$ and $\chi^2_{T_S_COV}$ are favored for detecting a misspecified between-covariance structure than other fit indices. As for a misspecified between-mean structure, $RMSEA_{PS_B}$, CFI_{PS_B} , and TLI_{PS_B} are recommended.

Even though previous studies indicated that some level-specific and target-specific model fit indices are recommended to detect misspecification in different levels or structures of MLGM, all previous studies only simulated MLGM with balanced data. Each simulation study also only simulated one type of trajectory. Even though some level-specific and target-specific model fit indices perform well in MLGM with balanced design with one type of trajectory, the influence of unbalanced datasets and other growth

trajectories are needed to be modeled to provide researchers more information about fit indices' performance. As an extension of the previous studies, this research investigated performance of level-specific and target-specific model fit indices under correctly specified MLGM with an unbalanced design. Besides, to assist empirical researchers, three different trajectories: accelerating, decelerating, and constant, were simulated. The goal was to examine the performance of level-specific and target-specific model fit indices and to determine which indices are not influence by unbalanced design and trajectory.

CHAPTER 3

METHOD

This chapter details methodology that used to meet the goals of the study. The study used a Monte Carlo simulation framework to examine a Multilevel Latent Growth Model (MLGM) framework. Conditions of different number of groups, unbalanced design, and trajectories were manipulated. The goal of the study is to inform researchers about the performance of different level-specific and target-specific model fit indices in correctly specified MLGM using a population design with five time points, an unbalanced design and different rates trajectories. The research questions to be investigated using the correctly specified MLGM framework are restated:

- (1) How are level-specific and target-specific fit indices impacted by sampling error, unbalanced design, and different growth trajectories?
- (2) Do the level-specific and target-specific fit indices demonstrate reasonable sensitivity to sampling error, unbalanced design, and different trajectories?

This chapter is divided into several parts. The first part discusses the Monte Carlo simulation and details the population model and the characteristics to be manipulated.

Second, this chapter describes framework of the Multilevel Latent Growth Model (MLGM), the different model fit indices used for evaluating the goodness fit of the MLGM, and the procedure for creating the level-specific and target-specific model fit indices. The third part describes the experimental procedures for generating the simulated data with conditions. The third section also illustrates how to analyze the level-specified and target-specified model indices to determine these fit indices' performance among different cluster sizes, unbalanced sample sizes, and trajectories.

Monte Carlo Simulation

The current research used the Monte Carlo framework (i.e., simulation). Monte Carlo studies are performed to examine the “best practices” and create “rules of thumb” for statistical models (Muthén, 1997). These simulation studies are typically used to estimate the performance of statistical estimators under varying design conditions. In educational research, Monte Carlo simulations are commonly used to inform methodological practice (Miller et al, 2000; Moser et al, 2012). The procedure of Monte Carlo simulation includes generating under hypothesized modeling conditions, drawing samples, estimating models, and then analyzing standard errors and parameter values over different samples (Muthén & Muthén, 2004). The advantage of using a Monte Carlo simulation is that various realistic conditions may be simulated and used to investigate model estimation information with the ability to compare the results to known, “true”, values (Roth & Switzer, 1995). Using Monte Carlo simulation, researchers can control the design conditions under which the simulation is conducted and change conditions such as sample size and number of factor indicators in different models.

The strategy for the current research was first to replicate a correctly specified MLGM with five time points and then analyze the impact of different design factors on level-specific and target-specific model fit indices. A correctly specified MLGM was simulated to represent a variety of possible educational contexts (e.g., variation in class or school sizes, unbalanced class or school sizes, and variation in trajectory). Outcomes for this study include which fit indices are the best, most consistent, and least sensitive under correctly specified MLGM with different design conditions.

Simulation and Analysis

Population model for data Generation

A Monte Carlo study was performed to evaluate the performance of both l-s fit indices ($\chi^2_{PS_B}$, $RMSEA_{PS_B}$, CFI_{PS_B} , TLI_{PS_B} , $SRMR_B$, $\chi^2_{PS_W}$, $RMSEA_{PS_W}$, CFI_{PS_W} , TLI_{PS_W} , $SRMR_W$) and t-s fit indices ($\chi^2_{T_S_COV}$, $RMSEA_{T_S_COV}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, $SRMR_{T_S_COV}$, $\chi^2_{T_S_MEAN}$, $RMSEA_{T_S_Mean}$, $CFI_{T_S_Mean}$, $TLI_{T_S_Mean}$, and $SRMR_{T_S_Mean}$) in a two-level correctly specified MLGM. The design factors include the number of groups, unbalanced group sizes, and different growth trajectories (constant, decelerating, accelerating).

Based on previous research, parameter settings from the LSAY (Longitudinal Study of American Youth) was used to simulate the correctly specified MLGM model (Hsu, 2019). The LSAY is a widely used exam to study the growth of mathematics and science performance (Ma & Wilkins, 2007). The parameters used for the population model are based on one MLGM study of LSAY, which contains 3,102 students from grade 7 to grade 11 nested within 52 schools (Hsu, 2019). The intraclass correlation

coefficients (ICCs) of five-time points ranged from .15 to .19. The ICC indicates the ratio between cluster-level variance and total variance (Jordan et al, 2009), and these values of ICC shows that cluster-level should not be ignored in this MLGM study (Lane, 2008).

In line with previous MLGM simulation studies (Palardy, 2008), a five-wave MLGM model was measured in this research. The five-time points, denoted as V1–V5, were assumed to be continuous data distributed on the standardized scale (i.e., Mean= 0 and SD = 1).

A two-level cluster data: students, and class, was considered. The parameters IB, LB, and QB indicate the intercept, linear growth parameters, and quadratic growth parameters for the between-level model. The IW, LW, and QW indicate the intercept, linear growth parameters, and quadratic growth parameters for the within-level model. For the between level model, the means of the factors IB, LB, and QB are defined as 49.96, 4.32, and – 0.13, whereas for the within-level model, the means of IW, LW, and QW are fixed at zero (Muthén, 1997). In the between-level model, the parameter settings for the mean structure and covariance structure are presented in matrices α_B and Φ_B , and mean structure and covariance structure in within model are presented in matrices α_W and Φ_W .

$$\alpha_B = \begin{bmatrix} 49.96 \\ 4.32 \\ -0.13 \end{bmatrix}$$

$$\alpha_W = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

In Φ_B , the variances of IB (16.20), LB (0.61), and QB (0.02) are the diagonal values in the matrix. The covariances among the IB, LB, and QB are the nondiagonal values in the matrix, and the covariance between IB and QB and covariance between LB and QB are constrained to be zero. In Φ_W , the variances of IW (71.45), LW (14.76), and QW (0.70) are the diagonal values in the matrix. The covariances among the IW, LW, and QW are the nondiagonal values. According to Wu and West (2010) in the SEM framework, the general population quadratic model for the population does not consider the covariance between the intercept and slope and covariance between linear and slope. In this way, we set the answer increasingly complex substantive questions for a longitudinal study for I and Q, and L and Q at both between-level and within-level be zero for simplicity.

$$\Phi_B = \begin{bmatrix} 16.2 & 2.82 & 0 \\ 2.82 & 0.61 & 0 \\ 0 & 0 & 0.02 \end{bmatrix}$$

$$\Phi_w = \begin{bmatrix} 71.45 & 6.76 & 0 \\ 6.76 & 14.76 & 0 \\ 0 & 0 & 0.070 \end{bmatrix}$$

The error variances for five-time points of between level model are set to 11.91, 15.25, 10.32, 12.59, and 1.93 and are uncorrelated over time. The error variances for five-time points of within level model are set to 1.80, 1.28, 0.06, 0.54, and 0.31, and these scores are also uncorrelated over time.

Three different design factors are simulated for the population model, including number of groups (NG), unbalanced groups size (GS), and different trajectories. Each

design factor contains three conditions. In this way, we have 27 total simulation conditions. Based on the recommendation that 1000-5000 replication is required to produce a stable result in Monte Carlo studies (Preacher et al., 2008), 1000 complete datasets based on population model were generated for each simulation condition. SAS 9.4 was used to simulate the datasets (Jiang, 2014). Part of SAS code used for generating data in this study are included in Appendix A. In the SAS steps, unbalanced group sizes were created firstly. In each group size, matrix and data for observation-level were simulated. Based on the matrix and data of observation-level, the matrix and data of group-level were simulated. The SAS steps were run for 1000 time for each condition.

The estimation of all the population models was carried out in Mplus 7.11 (Muthén & Muthén, 2017), using maximum likelihood estimation with robust standard errors (ESTIMATOR = MLR). The maximum number of iterations were set to 100 (ITERATIONS = 100) with 95 convergence criterion set to .000001 (CONVERGENCE = .000001). MLR are robust to non-normality and non-independence of observations when used with TYPE=COMPLEX (Muthén & Muthén, 2017). Our simulated datasets contain small sample sizes, which were non-normal samples. The students of simulated datasets are nested within each cluster, meaning the datasets are non-independence. MLR was the appropriate estimator for Mplus.

Design Conditions

NG and unbalanced GS

NG conditions were based on Wu, Kwok, and Willson's (2015) studies, and set at 50, 100, and 200. To maximize the effect of imbalance, the group sizes were chosen to be as different as possible. The highest number 200 conforms to Tanaka's (1987)

recommended lower limit for achieving good maximum likelihood estimates with normal data. The lower values 50 and 100 have been chosen because, in empirical multilevel modeling, it is hard to collect data from as many as 200 groups (Schaffer & Yucel, 2002).

As with Hox, Maas, and Brinkhuis's (2001) simulation study and the regression rule of thumb for multilevel research, each predictor requires at least ten observations (Ingels et al, 2013). The averages of unbalanced CS conditions are manipulated into three levels, 10, 20, and 50. In each level, we employ two distinct group sizes, with exactly half the groups being small and the other half being large. The large group size is three times as large as the small group size. For unbalanced GS is 10, small size is 5, and large sample size is 15; For unbalanced GS is 20, small size is 10, and large sample size is 30; For unbalanced GS is 50, small size is 25, and large sample size is 75. These three levels of CS range are also consistent with two large-scale national educational databases: the Early Childhood Longitudinal Study (Youn et al., 2011) and the Early Childhood Longitudinal Study (Wang et al., 2015).

Three different trajectories

So far, the Monte Carlo studies that investigate model fit indices in the latent growth model (e.g., Hsu et al., 2019; Wu & West, 2010; Wu, West, & Hughes, 2008) have only examined the MLGM with accelerating growth rate. However, as shown in chapter two, many researchers have adopted MLGMs that can accommodate different growth patterns, and in most cases, the growth rate was constant, accelerating, or decelerating. Therefore, the datasets were simulated with three different growth trajectories: constant growth, decelerating growth, and accelerating growth. When estimating the model parameters, all these three trajectory situations can be

accommodated by the same MLGM. Because researchers may encounter different growth trajectory, examining this condition will increase the generalizability of the study.

Both quadratic and constant growth patterns were modeled to be the same for both the within and the between level models. For the constant growth rate, the factor loadings of both levels' intercept factors were fixed at 1.0, and loadings of the linear growth slope factors were set to 0, 1, 2, 3, and 4. The quadratic slope factors were fixed as 0 for the constant growth rate. For the accelerating growth pattern, intercept factors and linear growth slope factors are the same with constant growth, but the quadratic slope factors were specified with quadratic factor loadings set to 0, 1, 4, 9, and 16. The decelerating growth pattern, intercept factors, and linear growth slope factors are also the same with constant growth, but the quadratic slope factors were specified with quadratic factor loadings set to 0, -1, -4, -9, and -16.

For the constant growth rate, the factor loadings for between level Λ_B and within level Λ_W for T (e.g., five) time points are set as:

$$\Lambda_B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \end{bmatrix} \quad \Lambda_W = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \end{bmatrix}$$

For the accelerating growth pattern, the factor loadings for between level Λ_B and within level Λ_W for T (e.g., five) time points are set as:

$$\Lambda_B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \quad \Lambda_W = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}$$

The decelerating growth pattern, the factor loadings for between level Λ_B and within level Λ_W for T (e.g., five) time points are set as:

$$\Lambda_B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & 2 & -4 \\ 1 & 3 & -9 \\ 1 & 4 & -16 \end{bmatrix} \quad \Lambda_W = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & 2 & -4 \\ 1 & 3 & -9 \\ 1 & 4 & -16 \end{bmatrix}$$

Data analysis and outcomes

This research includes two analyses. In the first analysis, we examined the descriptive information for model fit indices under different design factors: NG, unbalanced GS, or types of trajectories. The descriptive information includes the mean and standard deviation. For each trajectory, one table was generated to include all means and standard deviations of different NG and unbalanced GS by fit indices. In each table, the columns were different level-specific and target-specific model fit indices. The rows were different NG and unbalanced GS. Each fit index contained two rows: one row for mean and one for standard deviation. We also generated 5 box plots to show the distribution of model fit indices under design factors.

The second part of the analysis evaluated the sensitivity of both level-specific and target-specific model fit indices to different design factors. The results showed that which fit indices were practically significantly influenced by design factors. ANOVA with an individual model fit index's values as the dependent variables were conducted to evaluate influence of design factors. The ANOVA partitioned the total sum of squares of each fit index into different design factors. The total sum of squares of each fit index showed the variability of the values of each fit index across all replications under specific simulation

design factors. For ANOVA, we calculated the effect size, eta-squared (η^2), by dividing the Type III sum-of-square attributable to each design factor or the interaction between factors by the corrected total sum-of-square. η^2 describes the proportion of the variability accounted for a particular design factor or interaction effect term. In this study, each simulation condition has the same number of simulated datasets, resulting in orthogonal design factors. In this way, the Type III sum-of-squares from different factors were additive and non-overlapping, meaning the η^2 of each design factor could be calculated separately without considering other factors. Following Rogosa & Saner's (1992) study, we considered a moderate η^2 of .0588 to identify practically significant design factors for the fit indices' values. Note that when a fit index had a standard deviation close to 0, the impact of design factors on the fit index were trivial, even though the η^2 s were larger than .0588. As for our analysis, when fit indices have extremely low variability, we regarded design factors do not affect the fit indices. One table was generated for ANOVA results. In the table, the columns were different level-specific and target-specific model fit indices. The rows were: NG, unbalanced GS, interaction of NG and unbalanced GS, and trajectory. Each cell contains the value of η^2 .

Model Fit Indices for MLGM

Typical SEM Model Fit Indices

Researchers evaluate MLGM addressing two main issues: (a) evaluating the goodness of fit of the hypothesized model to the collected data, and (b) estimating parameters in the hypothesized model. Goodness of fit examines the extent to which the hypothesized models, proposed based on previous theories or findings, represent the relationships among the observed variables (DiStefano & Hess, 2005; DiStefano, 2016).

A poor-fitting model indicates that the hypothesized model widely differs from the underlying structure observed in the data and the model does not adequately summarize the relationships between the variables (Prosser, 1991; Hirvonen et al., 2012).

Methods to evaluate the goodness of fit of MLGMs has not been thoroughly studied, nor even well defined. One common approach when estimating MLGMs is to use global fit indices. There are five widely been proposed and widely applied to evaluate the model fit in MLGM: Chi-Square value (χ^2), the standardized root mean square residual (SRMR), comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA) and (Dalton & Schulz, 2006; Curran, 2003).

Chi-Square

The Chi-Square evaluates the amount of discrepancy between the covariances matrices of the data and the proposed model.

$$\chi^2 = \{-1/2(n-1)[tr(S\Sigma^{-1}) + \log|\Sigma| - \log|S| - p]\} = (n-1) * F$$

This formula shows the calculation of χ^2 with $\{1/2 p (p + 1)\} - t$ degree of freedom (df) in large samples. The df refers to the number of values involved in the calculations that vary or are freely estimated in the model. The number of parameters to estimate can range from few parameters (most values fixed, such in an LGM) to estimating all possible parameters, (i.e., saturated model). Here, p indicates the number of observed variables, and t symbolizes the number of estimated independent parameters. S describes the unrestricted sample covariance matrix, whereas $\Sigma(\theta)$ describes the restricted

covariance matrix. With the larger value of χ^2 , the hypothesized model is closer to the “true” model (Sirin, 2005; Linda et al, 1993).

χ^2 is commonly used because it is easier to compute than other model fit indices. It can also be used with categorical data and to check the if there is a “difference” between different groups of participants. There are drawbacks with the use of χ^2 as a fit index. From the above formula, we can know that the χ^2 test statistic is sensitive to sample size (Gerbing & Anderson, 1992). With an increasing sample size and a fixed number of df (meaning, XYZ), the χ^2 value increases and is not able to discriminate between good fitting models and poor fitting models (Wayman, 2003). The χ^2 may also causes a problem due to a lack of power, in that that a plausible model with a small sample size might be rejected (Zullig et al, 2014). Besides the sample size problem, χ^2 has other problems that limit its usefulness. For example, under ML-based (and other normal theory model estimation), because the χ^2 assumes multivariate normality, deviations from normality may result in poor χ^2 value even though the model is appropriately specified (Goos et al, 2013).

SRMR

Standardized Root Mean Square Residual (SRMR) is calculated based on the average of standardized residuals between the observed model's covariance matrices and the hypothesized model (Cohen et al, 2013). One of the reasons why SRMR is widely used in SEM studies as its relative independent from sample size (Cohen et al, 2013). The formula is shown as follows:

$$SRMR = \sqrt{(\sum_{i=1}^p \sum_{j=1}^i [\frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ii}s_{jj}}]^2) / p(p+1)/2}$$

where s_{ij} describes part of sample covariance matrix, and $\hat{\sigma}_{ij}$ represents part of covariance matrix for hypothesized model. P indicates the total number of observed variables (Chen, 2007; Carlson et al, 2009). Values for the SRMR ranges from zero to one. For well-fitted models, cut off values are supposed to be less than .05, and values as high as 0.08 are sometimes also deemed acceptable (Chou et al, 1998; Bentler & Dudgeon, 1996). However, a lower SRMR value does not always indicate a perfect model fit. When there are many parameters in the model and large sample sizes, SRMR also gives acceptable values even though the hypothesized model does not fit the dataset (Boulton, 2011).

RMSEA

The Root Mean Square Error of Approximation (RMSEA) evaluates the difference between the covariance matrix per degree of freedom of the data and the model's hypothesized covariance matrix (Chen, 2007). The RMSEA indicates the extent that hypothesized the model would fit the data covariance matrix (McArdle & Epstein, 1987). The formula is shown as follows:

$$RMSEA = \sqrt{\max\left(\left[\left(\chi^2/df\right) - 1\right]/(N - 1), 0\right)}$$

Here χ^2 describes chi-square value. df is degrees of freedom and N is the sample size (Boulton, 2011). The cut off value for RMSEA is smaller than 0.05, the model can be said to indicate that the model fits the analyzed data.

Unlike χ^2 and SRMR, RMSEA is not affected by the sample size, which means that RMSEA can still evaluate the model with small sample sizes (Clarke et al., 2008). Because RMSEA favors parsimony, it will choose the model with fewer estimated

parameters. With RMSEA, researchers could calculate the confidence interval (CI) around the value (Bandalos & Leite, 2013). A CI provides range around a point estimate that conveys the precision of the measurement. The CI associated with RMSEA tells researchers the possible range around the estimate. Typically, a 90% CI is estimated there is a 90% probability that the CI will contain the true value.

TLI

The Tucker-Lewis Index (TLI), also known as the Non-Normed Fit Index (NNFI), was developed against the Normed Fit Index (NFI) disadvantage regarding being affected by sample size. NFI analyzes the discrepancy between the chi-squared value of the hypothesized model and the chi-squared value of the null model. However, NFI tends to be negatively biased when the sample size is small. The TLI resolves some of these issues of negative bias. The TLI depends on the average size of the correlations; for example, if the average correlation between variables is high, then the TLI will be very high. Even though TLI is not affected significantly by the sample size, the TLI value can show poor fit when other fit indices are pointing towards good fit in models where small samples are used (Bentler, 1990; Boulton, 2011). TLI is calculated based on the below formula (Asparouhov, 2011).

$$TLI = \frac{(\chi_i^2/v_i) - (\chi_t^2/v_t)}{(\chi_i^2/v_i) - 1} = \frac{(F_i/v_i) - (F_t/v_t)}{(F_i/v_i) - (1/(n-1))}$$

v_i and v_t are the numbers of degrees of freedom for the saturated and hypothesized models, respectively. F is the value of the minimum fit function, and n describes sample size. Higher TLI values indicate better model-data fit, and 0.97 is accepted as the cut-off

value in most research studies (DiStefano, 2013). Because of TLI's non-normed nature, values can go above 1.0, making the value challenging to interpret (Graham et al, 2012).

CFI

The Comparative Fit Index (CFI) measures the extent to which the tested model is superior to the alternative model established with the manifest covariance matrix (Chen, 2007). CFI regards that all latent variables are uncorrelated and compares the covariance matrix of the sample with the null model's covariance. The formula is shown below:

$$CFI = 1 - \frac{\max [(\chi^2_t - \nu_t), 0]}{\max [(\chi^2_t - \nu_t), (\chi^2_i - \nu_i), 0]}$$

Where ν_i and ν_t are the degrees of freedom of the saturated model and the hypothesized model, respectively (Hsu et al, 2015), the CFI generates values between 0 and 1, and an acceptable cut off value for fit is larger than 0.95 (Hsu et al, 2015). This index is relatively independent of sample size and yields better results for studies with a small sample size (Chen et al, 2012). Because of this advantage, CFI is included in all SEM software as a critical reported model fit index (Hsu et al, 2015).

Researchers have relied heavily on above five global fit indices and the cutoff values to determine model-data goodness fit and model results of MLGM. (e.g., Enders, 2006; Graham & Coffman, 2012). However, previous studies have shown that global fit indices can only reveal the models' model fit but fail to detect if hypothesized the between-level model fit the data (Hsu, 2015). As an MLGM contains both between-level and within-level models and the sample size is much larger at the within-level than at the between-level, the global fit indices are likely to be dominated by the within-level, as shown in formulas above (Kaplan, 2009; Kwon, 2011). When global fit indices indicate

an excellent fitting model, a researcher does not know if both the between-level and within-level model fit well or because the global fit indices fail to detect the misspecification at the between-level (Leite & Stapleton, 2011). Besides, even though global fit indices output if the overall fit of a hypothesized model to a set of data, researcher does not know if the misspecification is due to the covariance matrice(s) or mean structure of MLGM. In this way, the results from global fit indices do not provide researchers information about how to modify the hypothesized MLGM model.

Level-specific Model Fit Indices for MLGM

Level-specific fit indices have been suggested to fix the disadvantages of global fit indices and evaluate the within-level model and the between-level model of MLGM separately (Hsu et al, 2015; Pornprasertmanit et al, 2013). SRMR for between level (SRMR_B) and SRMR for within level (SRMR_W) were the only fit indices that could be obtained from MLGM results to evaluate the between-level model and within-level in some commonly used statistical packages (e.g., Mplus). Other level-specific fit indices are not available from the commonly used SEM statistical packages and require fitting a specific model to obtain.

The general method to compute the between-level-specific fit (b-l-s) is to define the partially saturated model method (PS method). This method specifies the hypothesized MLGM model with hypothesized between-level model and with a saturated within-level model (Ryu & West, 2009). A saturated within-level model means the within-level model has zero degrees of freedom and a χ^2 test statistic equal to zero. In the same way, within-level-specific (w-l-s) fit can also be derived by using the PS method, which specifies the hypothesized within-group model along with a saturated between-

group model. The b-l-s can reflect the degree of misfit for the between level model, and the w-l-s able to reflect the degree of misfit for the hypothesized within-level model.

By using the PS method, b-l-s χ^2 test statistics ($\chi^2_{PS_B}$) and w-l-s χ^2 test statistics ($\chi^2_{PS_W}$) can directly be obtained from software output. The b-l-s χ^2 test statistics only reflect the model fit of the hypothesized between model and w-l-s χ^2 test statistics just indicate the model fit of the hypothesized within-level model (Schermerle-Engel et al, 2014). Other model fit indices are calculated based on different χ^2 test statistics. B-l-s fit indices, including RMSEA_{PS_B}, CFI_{PS_B}, and TLI_{PS_B}, can be calculated after $\chi^2_{PS_B}$ is obtained, and w-l-s fit indices, including RMSEA_{PS_W}, CFI_{PS_W}, and TLI_{PS_W} can be computed after $\chi^2_{PS_W}$ is obtained.

Target-specific Model Fit Indices for MLGM

Target-specific (t-s) fit indices will also be calculated. Traditional fit indices can evaluate the overall fit of a model, but the indices fail to specify either mean structure or covariance structure of hypothesized MLGM do not fit the data. Thus, t-s fit indices could provide further information to researchers about how to modify a model. T-s fit evaluate the fit of the covariance or mean structure and will tell us how to modify the covariance or mean structure of the MLGM. Target-specific fit indices for the mean (t-s-m) structure, such as $\chi^2_{T_S_Mean}$, SRMR_{T_S_Mean}, CFI_{T_S_Mean}, TLI_{T_S_Mean}, and RMSEA_{T_S_Mean}, can be calculated by saturating the covariance structure of the model. Target-specific fit indices for the covariance structure (t-s-cov), such as $\chi^2_{T_S_COV}$, SRMR_{T_S_COV}, CFI_{T_S_COV}, TLI_{T_S_COV}, RMSEA_{T_S_COV}, can be calculated by saturating the mean structure of the model (Wu & West, 2010). With the saturated mean structure, research can freely estimate the intercepts for all repeated measures and fixed the

different growth factors' means to zero. Based on recommendations from Wu and West (2010) and Ryu and West (2009), $\chi^2_{T_S_COV}$, $SRMR_{T_S_COV}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, and $RMSEA_{T_S_COV}$ were generated by saturating the within-level model and the mean structure of the between-level model. $SRMR_{T_S_Mean}$, $CFI_{T_S_Mean}$, $TLI_{T_S_Mean}$, $RMSEA_{T_S_Mean}$ and will be generated by saturating the within-level model and the covariance structure of the between-level model.

Calculation procedures for level-specific and target-specific model fit indices for MLGM

Chi-square

Based on Schumacker and Lomax (2010), b-l-s chi-square ($\chi^2_{PS_B}$) can be calculated with the equation:

$$\chi^2_{PS_B} = F_{ML} [\sum_B (\hat{\theta}), \sum_W (\hat{\theta}s)] - F_{ML} [\sum_B (\hat{\theta}s), \sum_W (\hat{\theta}s)],$$

In the equation, $F_{ML} [\sum_B (\hat{\theta}), \sum_W (\hat{\theta}s)]$ is the value of the fitting function for a two-level model. The model contains a hypothesized model for between level and a identified model for within level. $F_{ML} [\sum_B (\hat{\theta}s), \sum_W (\hat{\theta}s)]$ indicates the fit function for a two-level model, which contains both the saturated within-level and between-level models. The degrees of freedom, df_{PS_B} , is the difference in the number of parameters in the model with just saturated within-level model and model with both the between-level and within-level models saturated.

Similarly, A w-l-s chi-square ($\chi^2_{PS_W}$) is computed by:

$$\chi^2_{PS_W} = F_{ML} [\sum_B (\hat{\theta}s), \sum_W (\hat{\theta})] - F_{ML} [\sum_B (\hat{\theta}s), \sum_W (\hat{\theta}s)].$$

In the equation, $F_{ML} [\sum_B (\hat{\theta}s), \sum_W (\hat{\theta})]$ describes the value of the fit function for the two-level model with the hypothesized within-level model and the saturated between-level model. The degrees of freedom (df_{PS_W}) is the difference between the parameters in the model with just saturated between-level model and model with both saturate models. T-s-m chi-square ($\chi^2_{T_S_MEAN}$) and t-s-cov chi-square ($\chi^2_{T_S_COV}$) use same equation with different $F_{ML} [\sum_B (\hat{\theta}), \sum_W (\hat{\theta}s)]$.

To calculate $\chi^2_{T_S_MEAN}$, researchers need to estimate the saturated covariance structure of a between level model. Wu and West (2010) indicated that a saturated between-covariance structure could be achieved by freely estimating the variances at the between level and constraining the variances of the factors of intercept, linear slope, and quadratic slope to be zero. The degrees of freedom ($df_{T_S_MEAN}$) is the difference in the number of parameters in the model saturated between the covariance structure model and model with both between and within saturated models. $F_{ML} [\sum_B (\hat{\theta}), \sum_W (\hat{\theta}s)]$ for $\chi^2_{T_S_COV}$ needs to saturate the mean structure of between-level model.

Wu and West (2010) also indicate that saturated between-mean structure can be calculated by freely estimating the intercepts of items at the between-level and constraining the intercepts of the factors of the intercept, linear slope, and quadratic slope to be zero. The degrees of freedom ($df_{T_S_COV}$) is the difference between the parameters model numbers with saturated between-mean structure and fully saturated model.

RMSEA

The b-l-s RMSEA ($RMSEA_{PS_B}$) is calculated based on $\chi^2_{PS_B}$ and its corresponding degree of freedom. The equation is as follow:

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

In the equation, J stands for the number of groups for the between-level, which is the sample size for between level. The number of groups is also regarded as a penalty for a large sample size. If the $\chi^2_{PS_B}$ is smaller than df_{PS_B} , $RMSEA_{PS_B}$ will be set as zero.

W-l-s RMSEA ($RMSEA_{PS_W}$) can be calculated from the following equation, with N as the total sample size for within-level.

$$RMSEA_{PS_W} = \sqrt{Max(\frac{\chi^2_{PS_W} - df_{PS_W}}{df_{PS_W}(N)}, 0)}.$$

T-s-m RMSEA ($RMSEA_{T_S_MEAN}$) and t-s-cov RMSEA ($RMSEA_{T_S_COV}$) can also be computed through the same equations just by using different degree of freedom:

$\chi^2_{T_S_MEAN}$ with $df_{T_S_MEAN}$ and $\chi^2_{T_S_COV}$ with $df_{T_S_COV}$.

CFI

The b-l-s CFI (CFI_{PS_B}) is calculated as:

$$CFI_{PS_B} = 1 - \frac{Max[(\chi^2_{PS_B} - df_{PS_B}), 0]}{Max[(\chi^2_{I_B, S_W} - df_{I_B, S_W}), 0]},$$

where $\chi^2_{I_B, S_W}$ represents the fit of a model with a saturated between-level model and a saturated within-level model. The saturated between-level model is a model in which only the mean of the intercept factor and residual variances are freely estimated. The formula for $\chi^2_{I_B, S_W}$ is:

$$\chi^2_{I_B, S_W} = F_{ML} [\sum_W (\hat{\theta}_I), \sum_W (\hat{\theta}_S)] - F_{ML} [\sum_B (\hat{\theta}_S), \sum_W (\hat{\theta}_S)]$$

The degrees of freedom (df_{l_B, s_w}) is the difference between the numbers of parameters of the fully saturated model and the model with the saturated between-level model and a saturated within-level model.

The w-l-s CFI (CFI_{ps_w}) can be computed by

$$CFI_{ps_w} = 1 - \frac{\text{Max}[(\chi^2_{ps_w} - df_{ps_w}), 0]}{\text{Max}[(\chi^2_{s_B, l_w} - df_{s_B, l_w}), 0]}.$$

$\chi^2_{s_B, l_w}$ represents the fit for a two-level model, which has a saturated between-level model and a saturated within-level model. The saturated within-level model shows an intercept-only growth model, and residual variances are freely estimated. The degrees of freedom (df_{s_B, l_w}) are the difference in the numbers of parameters in the fully saturated model and the model with the saturated within-level model. $\chi^2_{s_B, l_w}$ represents the fit of a model with a saturated within-level model and a saturated between-level model:

$$\chi^2_{s_B, l_w} = F_{ML} [\sum_B (\hat{\theta}_s), \sum_W (\hat{\theta}_l)] - F_{ML} [\sum_B (\hat{\theta}_s), \sum_W (\hat{\theta}_s)]$$

T-s-m CFI (CFI_{T_S_MEAN}) and t-s-cov CFI (CFI_{T_S_COV}) were calculated through same equation by substituting $\chi^2_{ps_B}$ (and df_{ps_B}) with $\chi^2_{T_S_MEAN}$ (and $df_{T_S_MEAN}$) or $\chi^2_{T_S_COV}$ (and $df_{T_S_COV}$).

TLI

The b-l-s TLI (TLI_{ps_B}) computed by comparing the hypothesized between-level model and the saturated between-level model given that the within-level model is saturated. Computing TLI_{ps_B} uses similar information for computing CFI_{ps_B}:

$$TLI_{PS_B} = \frac{\frac{\chi^2_{I_B,S_W}}{df_{I_B,S_W}} - \frac{\chi^2_{PS_B}}{df_{PS_W}}}{\frac{\chi^2_{I_B,S_W}}{df_{I_B,S_W}} - 1}.$$

The w-l-s TLI (TLI_{PS_W}) is calculated by comparing the hypothesized within-level model and the saturated within-level model given that the between-level model is saturated. The equation for TLI_{PS_W} is presented as follow:

$$TLI_{PS_W} = \frac{\frac{\chi^2_{I_B,S_W}}{df_{I_B,S_W}} - \frac{\chi^2_{PS_B}}{df_{PS_W}}}{\frac{\chi^2_{I_B,S_W}}{df_{I_B,S_W}} - 1}.$$

T-s-m TLI ($TLI_{T_S_MEAN}$) and t-s-cov TLI ($TLI_{T_S_COV}$) are also calculated with similar information for computing $CFI_{T_S_COV}$ and $CFI_{T_S_MEAN}$ by substituting $\chi^2_{PS_B}$ (and df_{PS_B}) with $\chi^2_{T_S_MEAN}$ (and $df_{T_S_MEAN}$) or $\chi^2_{T_S_COV}$ (and $df_{T_S_COV}$).

SRMR

The b-l-s SRMR ($SRMR_B$) and w-l-s SRMR ($SRMR_W$) models are reported directly by the software program (i.e., Mplus) to be used in the present study. T-s-m SRMR ($SRMR_{T_S_MEAN}$) can also be derived from software program, using the model-implied variance matrix, which is generated by a multilevel model with a saturated between-covariance structure combined with a saturated within-level model. Similarly, T-s-cov SRMR ($SRMR_{T_S_COV}$) can also be calculated with the software program, using a multilevel model with a saturated between-mean structure and a saturated within-level model.

Summary

We evaluated the descriptive information of the level-specific and target-specific model fit indices under various simulation conditions in the first set of analysis. The second set of analyses targeted at evaluating the extent to which the fit indices could be influenced by sampling errors, unbalanced design, and trajectories, based on simulated data with different conditions derived from a correctly specified MLGM. Our simulation design included three levels of NG (50, 100, and 200), three levels of unbalanced GS (5/15, 10/20, and 25/75), and three trajectories (accelerating, decelerating, and linear). ANOVA results, in terms of η^2 , presented the influences of NG, unbalanced GS, and three trajectories. In the first set of analyses, we expect the means values of all fit indices are within the range of cut off values. The standard deviations are expected to be as small. In the second set of analyses, we expect η^2 of all ANOVA results to be smaller than 0.0588. The smaller η^2 s show that different fit indices to be less influenced by sampling errors arising from a small sample size and unbalanced design and less influenced by trajectories. Based on the results of two analyses, we made recommendations for practical and theoretical research about fit indices.

CHAPTER 4

RESULTS

In this chapter, we evaluate the results of the two analyses of data simulated under a correctly specified MLGM under various conditions. The first set of analyses examined means and standard deviations of the level-specific and target-specific χ^2 test statistics and model fit indices. The second set of analyses evaluated whether different level-specific and target-specific χ^2 test statistics and fit indices of interest were robust to sampling errors, unbalanced design, and different change trajectories. As this was one of the first investigations of MLGM with unbalanced group sizes, the trajectories at both the between and within different levels were set to be equal. Traditional cutoff criteria of the fit indices used with typical SEM studies (e.g., RMSEA- $<.06$; CFI and TLI $>.95$; SRMR- $<.08$; Hu & Bentler, 1999) were examined with multilevel models to determine if these recommended levels were able to accurately identify correct models across different number of groups, different unbalanced group sizes, and different trajectories.

To summarize results for the first analysis, one table of fit indices descriptive statistics was generated for each trajectory. Each table included means and standard deviations of model fit indices under all simulation conditions. As there were three levels of NG (50, 100, and 200) and three levels of unbalanced GS (5/15, 10/20, and 25/75), there were 9 simulation conditions included in each table.

The second analysis conducted an ANOVA by each model fit index as the dependent variable to evaluate influence of the design factors. For the ANOVA, we calculated the effect size, eta-squared (η^2). Following Bryk and Raudenbush's (1992) study, when the η^2 value was larger than .0588, the design factor was regarded as having an effect on the fit indices. Box plots were generated for fit indices with η^2 s larger than .0588 to show the variability of model fit indices under design factors.

Convergence Rates

Under the design conditions, the convergence rates over the 1,000 replications were 100% across all cells in the design. Thus, even under the smallest sample size (number of groups (NG) = 50, with an unbalanced group size (GS) = 5), the analysis was unlikely to encounter convergence problems. Results reported in this chapter were summarized across all replications.

Analysis 1: Descriptive Review of Conditions

Accelerating Growth Trajectory. Table 4.1 summarized means and standard deviations of all fit indices under accelerating growth trajectory over the within each design cell. As there were three levels of NG (50, 100, and 200) and three levels of unbalanced GS (5/15, 10/20, and 25/75), there were 9 simulation conditions included in Table 4.1.

Table 4.1 *Descriptive statistics of model fit indices by NG and unbalanced GS for the accelerating growth trajectory*

Fit Index	NG	50	50	50	100	100	100	200	200	200
	GS	5/15	10/30	25/75	5/15	10/30	25/75	5/15	10/30	25/75
χ^2 Test Statistics										
$\chi^2_{PS_B}$	Mean	29.38	14.76	5.13	11.08	5.16	4.37	5.30	4.31	4.01
	SD	240.63	68.62	8.89	46.92	4.42	3.14	4.85	3.01	2.83
$\chi^2_{PS_W}$	Mean	14.16	16.31	4.31	14.40	4.62	4.03	4.92	4.17	3.96
	SD	32.34	164.39	3.34	93.74	3.77	2.91	3.94	3.03	3.01
$\chi^2_{T_S_COV}$	Mean	137.53	100.67	73.21	87.20	102.43	131.03	121.63	179.80	249.44
	SD	729.46	532.83	23.19	58.96	38.77	28.76	39.91	36.79	39.47
$\chi^2_{T_S_Mean}$	Mean	80.33	26.39	23.57	22.75	25.71	35.74	27.49	40.92	62.85
	SD	828.27	62.94	9.691	19.68	11.06	10.92	12.00	12.18	14.44
RMSEA-related fit indices										
RMSEA _{PS_B}	Mean	0.054	0.027	0.008	0.027	0.010	0.005	0.010	0.005	0.003
	SD	0.100	0.046	0.012	0.035	0.012	0.007	0.013	0.007	0.004
RMSEA _{PS_W}	Mean	0.045	0.024	0.007	0.027	0.008	0.004	0.009	0.005	0.003
	SD	0.058	0.052	0.009	0.045	0.011	0.006	0.012	0.007	0.004
RMSEA _{T_S_COV}	Mean	0.144	0.098	0.061	0.102	0.081	0.059	0.089	0.078	0.059
	SD	0.128	0.061	0.010	0.032	0.014	0.007	0.015	0.008	0.005
RMSEA _{T_S_Mean}	Mean	0.074	0.044	0.029	0.042	0.035	0.028	0.037	0.034	0.028
	SD	0.124	0.029	0.009	0.022	0.011	0.005	0.011	0.006	0.004
CFI-related fit indices										
CFI _{PS_B}	Mean	0.995	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	SD	0.035	0.008	0.000	0.005	0.000	0.000	0.000	0.000	0.000
CFI _{PS_W}	Mean	0.998	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	SD	0.007	0.016	0.000	0.010	0.000	0.000	0.000	0.000	0.000
CFI _{T_S_COV}	Mean	0.977	0.990	0.996	0.990	0.994	0.997	0.993	0.994	0.997
	SD	0.061	0.033	0.001	0.006	0.002	0.001	0.002	0.001	0.001
CFI _{T_S_Mean}	Mean	0.991	0.998	0.999	0.998	0.999	0.999	0.999	0.999	0.999
	SD	0.061	0.008	0.000	0.002	0.001	0.000	0.001	0.000	0.000
TLI-related fit indices										
TLI _{PS_B}	Mean	0.972	0.994	1.000	0.996	1.000	1.000	1.000	1.000	1.000

TLI _{PS_w}	SD	0.254	0.042	0.002	0.027	0.001	0.000	0.002	0.001	0.000
	Mean	0.989	0.994	1.000	0.994	1.000	1.000	1.000	1.000	1.000
TLI _{T_S_COV}	SD	0.035	0.079	0.001	0.051	0.001	0.000	0.001	0.001	0.000
	Mean	0.924	0.970	0.990	0.973	0.982	0.990	0.979	0.983	0.990
TLI _{T_S_Mean}	SD	0.403	0.147	0.003	0.018	0.006	0.002	0.007	0.003	0.002
	Mean	0.959	0.994	0.997	0.995	0.996	0.998	0.996	0.997	0.998
	SD	0.460	0.022	0.001	0.006	0.002	0.001	0.002	0.001	0.001
SRMR-related fit indices										
SRMR _B	Mean	0.012	0.008	0.005	0.008	0.005	0.004	0.005	0.003	0.002
	SD	0.009	0.005	0.003	0.005	0.003	0.002	0.003	0.002	0.001
SRMR _W	Mean	0.004	0.002	0.001	0.002	0.002	0.001	0.002	0.001	0.001
	SD	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
SRMR _{T_S_COV}	Mean	0.172	0.183	0.193	0.170	0.184	0.193	0.168	0.183	0.193
	SD	0.042	0.027	0.019	0.029	0.019	0.013	0.021	0.012	0.009
SRMR _{T_S_Mean}	Mean	0.067	0.070	0.072	0.066	0.070	0.071	0.068	0.072	0.072
	SD	0.035	0.030	0.021	0.028	0.021	0.015	0.021	0.014	0.011

70 RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker–Lewis index. SRMR = standardized root mean square residual. Subscripted PS = partially saturated model method. Subscripted TS = target-specific fit indices. Subscripted B = between-level model. Subscripted W = within-level model. Subscripted COV = fit index for evaluating between-covariance structure. Subscripted MEAN = fit index for evaluating between-mean structure.

χ^2 test statistics: Accelerating Trajectory

In Table 4.1, means and standard deviations for the different χ^2 test statistics, $\chi^2_{PS_B}$, $\chi^2_{PS_W}$, $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, were calculated. When NG increased from 50 to 200 and the unbalanced GS increased from 5/15 to 25/75, values of χ^2 decreased, with $\chi^2_{PS_B}$ mean values ranged from 29.38 to 4.01 and $\chi^2_{PS_W}$ mean values ranged from 14.16 to 3.96. Also, the standard deviation of $\chi^2_{PS_B}$ decreased from 240.63 to 2.83 as NG and GS increased and standard deviation of $\chi^2_{PS_W}$ decrease from 164.37 to 3.01. Both indices showed that the average estimated $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ approached the expected value (i.e., model degrees of freedom of 4) when NG = 100 and unbalanced GS = 10/30.

For the $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, mean values did not approach acceptable model fit when total sample size increased. When NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, average fit indices decreased with $\chi^2_{T_S_COV}$ reporting mean values ranging from 73.21 to 249.44 and $\chi^2_{T_S_MEAN}$ reporting mean values ranging from 80.33 to 22.75. The standard deviation values decreased as total sample size increased; $\chi^2_{T_S_MEAN}$ decreased from 729.46 to 28.76 and $\chi^2_{T_S_COV}$ values decreased from 828.27 to 9.69. A total sample size over 1,000 was necessary for $\chi^2_{PS_W}$ and $\chi^2_{PS_B}$ to appropriately identify correct between-level and within-level models with an accelerating growth trajectory.

RMSEA-related fit indices: Accelerating Trajectory

In Table 4.1, means and standard deviations for the RMSEA-related fit indices, $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $RMSEA_{T_S_COV}$, and $RMSEA_{T_S_MEAN}$, were calculated. When NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, all $RMSEA_{PS_B}$ and $RMSEA_{PS_W}$ values were indicative of good model fit (i.e., <.06). The

standard deviation values were small and also decreased as NG and GS increased, with $RMSEA_{PS_B}$ decrease from 0.100 to 0.004 and standard deviation of $RMSEA_{PS_W}$ decrease from 0.058 to 0.004.

Means of $RMSEA_{T_S_COV}$ did not approach values indicative of acceptable model under all sample size combinations tested. For the $RMSEA_{T_S_MEAN}$, the data analysis showed that means of $RMSEA_{T_S_MEAN}$ approached the values indicative of acceptable model, except at the smallest sample size (NG = 50, unbalanced GS = 5/15). The standard deviation of $RMSEA_{T_S_COV}$ decreased from 0.128 to 0.005 and standard deviation of $RMSEA_{T_S_MEAN}$ decreased from 0.124 to 0.004. Overall, $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, and $RMSEA_{T_S_MEAN}$ values suggested that the accelerating growth trajectory model fit the data under most design conditions.

CFI-related fit indices and TLI-related fit indices: Accelerating Trajectory

Means of CFI-related fit indices and TLI-related fit indices were indicative of good model fit (i.e., $>.95$) under all simulation conditions. There was only one mean value less than the cutoff denoting good model fit: when NG was 50 and unbalanced GS was set to 5/15, the mean of $TLI_{T_S_COV}$ was 0.924. Most standard deviations of CFI-related fit indices and TLI-related fit indices were nearly 0, except when the sample sizes were small.

SRMR-related fit indices: Accelerating Trajectory

Table 4.1 reports means and standard deviations for the SRMR-related fit indices, including $SRMR_B$, $SRMR_W$, $SRMR_{T_S_COV}$, and $SRMR_{T_S_MEAN}$. All $SRMR_B$, $SRMR_W$, and $SRMR_{T_S_MEAN}$ mean values illustrated good model fit (i.e., $<.08$); however, means of $SRMR_{T_S_COV}$ approached values of poor model fit (i.e., $<.08$) under all simulation

conditions. The standard deviations values of $SRMR_B$ and $SRMR_w$ were close to zero as the increase of sample sizes. The standard deviation of $SRMR_{T_S_COV}$ range from 0.009 to 0.042 and standard deviation of $SRMR_{T_S_MEAN}$ range from 0.011 to 0.035, showing little variability SRMR-related fit indices.

Decelerating Growth Trajectory. Table 4.2 summarizes the means and standard deviations of all fit indices under decelerating growth trajectory. The means and standard deviations of fit indices were calculated based 1000 replications within each simulation condition. As we have three levels of NG (50, 100, and 200) and three levels of unbalanced GS (5/15, 10/20, and 25/75), there were 9 simulation conditions included in Table 4.2.

χ^2 test statistics: Decelerating Trajectory

In Table 4.2, means and standard deviations for the different χ^2 test statistics, $\chi^2_{PS_B}$, $\chi^2_{PS_W}$, $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, were calculated. When NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, $\chi^2_{PS_B}$ reported mean values decreased as did the standard deviation values. The average $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ approached the expected value (i.e., 4 degrees of freedom) when NG = 50 and unbalanced GS = 10/30. For $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, mean values of $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ did not approach acceptable model fit when total sample size increased, but the standard deviations of $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$ did become smaller under these conditions. When NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, $\chi^2_{T_S_COV}$ reported mean values ranging from 249.75 to 72.72 and $\chi^2_{T_S_MEAN}$ reported mean values ranging from 63.75 to 22.74. The standard deviation of $\chi^2_{T_S_COV}$ decreased from 875.22 to 21.90 and standard deviation of $\chi^2_{T_S_MEAN}$ decreased from 329.02 to 9.28. A total

sample size over 500 was necessary for $\chi^2_{\text{PS}_w}$ and $\chi^2_{\text{PS}_B}$ to appropriately identify correct between-level and within-level models with all trajectories.

Table 4.2 Descriptive statistics of model fit indices by NG and unbalanced GS for the decelerating growth trajectory

Fit Index	NC	50	50	50	100	100	100	200	200	200
	CS	5/15	10/30	25/75	5/15	10/30	25/75	5/15	10/30	25/75
χ^2 Test Statistics										
$\chi^2_{PS_B}$	Mean	17.70	12.05	4.64	18.37	5.04	4.33	5.28	4.25	4.07
	SD	83.25	52.50	3.53	231.34	4.20	3.05	4.61	3.12	2.88
$\chi^2_{PS_W}$	Mean	24.51	14.40	4.22	10.17	4.77	4.10	5.06	4.19	4.01
	SD	210.09	172.11	2.971	40.96	5.20	2.86	4.53	3.08	2.86
$\chi^2_{T_S_COV}$	Mean	179.80	79.48	72.72	93.00	101.19	132.41	121.99	179.53	249.75
	SD	875.22	87.12	21.90	94.98	33.75	27.33	42.01	37.92	63.75
$\chi^2_{T_S_Mean}$	Mean	52.36	22.74	22.96	24.37	25.20	35.86	27.82	40.61	63.75
	SD	329.02	16.23	9.28	41.91	10.29	11.23	12.54	12.78	14.54
RMSEA-related fit indices										
RMSEA _{PS_B}	Mean	0.045	0.026	0.008	0.028	0.010	0.005	0.010	0.005	0.003
	SD	0.071	0.039	0.010	0.054	0.012	0.006	0.013	0.007	0.004
RMSEA _{PS_W}	Mean	0.048	0.024	0.007	0.025	0.009	0.004	0.010	0.005	0.003
	SD	0.091	0.047	0.009	0.033	0.012	0.006	0.013	0.007	0.004
RMSEA _{T_S_COV}	Mean	0.145	0.095	0.060	0.104	0.081	0.059	0.089	0.078	0.059
	SD	0.169	0.036	0.010	0.037	0.014	0.007	0.015	0.009	0.004
RMSEA _{T_S_Mean}	Mean	0.072	0.042	0.029	0.044	0.035	0.028	0.037	0.034	0.028
	SD	0.089	0.022	0.009	0.024	0.010	0.006	0.012	0.007	0.004
CFI-related fit indices										
CFI _{PS_B}	Mean	0.997	0.999	1.000	0.998	1.000	1.000	1.000	1.000	1.000
	SD	0.016	0.005	0.000	0.025	0.000	0.000	0.000	0.000	0.000
CFI _{PS_W}	Mean	0.996	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	SD	0.034	0.019	0.000	0.005	0.000	0.000	0.000	0.000	0.000
CFI _{T_S_COV}	Mean	0.978	0.991	0.997	0.990	0.994	0.997	0.993	0.994	0.997
	SD	0.062	0.009	0.001	0.009	0.002	0.990	0.003	0.001	0.000
CFI _{T_S_Mean}	Mean	0.992	0.998	0.999	0.998	0.999	0.999	0.999	0.999	0.999
	SD	0.048	0.002	0.000	0.004	0.001	0.000	0.001	0.000	0.000
TLI-related fit indices										
TLI _{PS_B}	Mean	0.985	0.995	1.000	0.992	1.000	1.000	1.000	1.000	1.000

TLI _{PS_W}	SD	0.079	0.027	0.001	0.124	0.001	0.000	0.001	0.001	0.000
	Mean	0.978	0.994	1.000	0.996	1.000	1.000	1.000	1.000	1.000
TLI _{T_S_COV}	SD	0.221	0.094	0.001	0.025	0.002	0.000	0.001	0.001	0.000
	Mean	0.893	0.975	0.990	0.971	0.982	0.990	0.979	0.983	0.990
TLI _{T_S_Mean}	SD	1.281	0.025	0.003	0.027	0.006	0.002	0.007	0.004	0.001
	Mean	0.975	0.995	0.998	0.994	0.997	0.998	0.996	0.997	0.998
	SD	0.167	0.005	0.001	0.012	0.002	0.001	0.002	0.001	0.001
SRMR-related fit indices										
SRMR _B	Mean	0.012	0.008	0.005	0.008	0.005	0.003	0.005	0.003	0.002
	SD	0.009	0.005	0.003	0.005	0.003	0.002	0.003	0.002	0.001
SRMR _W	Mean	0.003	0.002	0.001	0.002	0.002	0.001	0.002	0.001	0.001
	SD	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
SRMR _{T_S_COV}	Mean	0.173	0.182	0.193	0.168	0.184	0.193	0.167	0.184	0.193
	SD	0.043	0.026	0.019	0.028	0.019	0.013	0.021	0.013	0.009
SRMR _{T_S_Mean}	Mean	0.068	0.070	0.071	0.068	0.071	0.071	0.069	0.072	0.072
	SD	0.036	0.029	0.022	0.028	0.021	0.015	0.021	0.015	0.010

76 RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker–Lewis index. SRMR = standardized root mean square residual. Subscripted PS = partially saturated model method. Subscripted TS = target-specific fit indices. Subscripted B = between-level model. Subscripted W = within-level model. Subscripted COV = fit index for evaluating between-covariance structure. Subscripted MEAN = fit index for evaluating between-mean structure.

RMSEA-related fit indices: Decelerating Trajectory

In Table 4.2, means and standard deviations for the RMSEA-related fit indices, $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $RMSEA_{T_S_COV}$, and $RMSEA_{T_S_MEAN}$, were calculated. When NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, all mean values of $RMSEA_{PS_B}$ and $RMSEA_{PS_W}$ approached values indicative of good model fit (i.e., $<.06$) and again, standard deviation values decreased as both NG and GS increased. All mean values of $RMSEA_{T_S_COV}$ did not illustrate acceptable model under all tested sample sizes. For the $RMSEA_{T_S_MEAN}$, the analysis showed that means of $RMSEA_{T_S_MEAN}$ approached values indicative of acceptable model fit, except the smallest sample size (NG = 50, unbalanced GS = 5/15). The standard deviation of $RMSEA_{T_S_COV}$ and $RMSEA_{T_S_MEAN}$ decreased as sample size increased.

CFI-related fit indices and TLI-related fit indices: Decelerating Trajectory

In Table 4.2, means and standard deviations for the CFI-related fit indices and TLI-related fit indices were presented. As with the accelerating trajectory results, Means of the CFI-related fit indices and TLI-related fit indices were indicative of good model fit (i.e., $>.95$) under all simulation conditions. There was only one mean value less than the tested cutoff: when NG was 50 and unbalanced GS was 5/15, the mean of $TLI_{T_S_COV}$ was 0.893. Most standard deviations of CFI-related fit indices and TLI-related fit indices were nearly 0, except when the sample sizes were small. In summary, the means and small standard deviations show that all CFI-related fit indices and TLI-related fit indices suggest that the tested model with decelerating growth trajectory fits the data for the majority of conditions.

SRMR-related fit indices Decelerating Trajectory

Table 4.2 reports means and standard deviations for the SRMR-related fit indices (SRMR_B, SRMR_w, SRMR_{T_S_COV}, and SRMR_{T_S_MEAN}). All mean values of SRMR_B, SRMR_w, and SRMR_{T_S_MEAN} illustrated good model fit (i.e., <.08). However, means of SRMR_{T_S_COV} yielded larger values approaching poor model fit under all simulation conditions. The standard deviations values of SRMR_B and SRMR_w were close to a value 0. Values of of the target specific indices were larger, but the values were still low (e.g., SRMR_{T_S_COV} ranging from 0.009 to 0.043 and standard deviation of SRMR_{T_S_MEAN} ranging from 0.010 to 0.036). In general, SRMR_B, SRMR_w, and SRMR_{T_S_MEAN} were able to identify a correctly specified MLGM under most sample size scenarios.

Linear Growth Trajectory. Table 4.3 summarizes the means and standard deviations of all fit indices under a linear growth trajectory.

χ^2 test statistics: Linear Trajectory

Descriptive statistics in Table 4.3 showed that when NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, $\chi^2_{PS_B}$ and $\chi^2_{PS_w}$ mean values and standard deviation values decreased. Both indices showed that the average estimated $\chi^2_{PS_B}$ and $\chi^2_{PS_w}$ approached the expected value (i.e., 4 degrees of freedom) when NG = 50 and unbalanced GS = 25/75. A total sample size over 1250 was necessary for $\chi^2_{PS_w}$ and $\chi^2_{PS_B}$ to appropriately identify correct between-level and within-level models when a linear growth trajectory was modeled. For the $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$, the descriptive values mean of $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ did not approach acceptable model fit when total sample size increased.

Table 4.3 Descriptive statistics of model fit indices by NG and unbalanced GS for the linear growth trajectory

Fit Index	NC	50	50	50	100	100	100	200	200	200
	CS	5/15	10/30	25/75	5/15	10/30	25/75	5/15	10/30	25/75
χ^2 Test Statistics										
$\chi^2_{PS_B}$	Mean	26.12	14.76	5.13	11.09	5.16	4.38	5.28	4.25	7.38
	SD	129.19	68.63	8.89	46.92	4.43	3.14	4.61	3.12	3.91
$\chi^2_{PS_W}$	Mean	44.99	16.32	4.31	14.41	4.63	4.03	5.06	4.19	6.98
	SD	491.81	164.40	3.34	93.74	3.78	2.91	4.53	3.08	3.68
$\chi^2_{T_S_COV}$	Mean	585.53	100.68	73.22	87.20	102.43	131.04	121.99	179.53	766.68
	SD	753.26	532.84	23.20	58.96	38.78	28.76	42.01	37.92	102.48
$\chi^2_{T_S_Mean}$	Mean	36.73	26.40	23.57	22.76	25.72	35.74	27.82	40.61	114.34
	SD	51.96	62.95	9.69	19.69	11.07	10.93	12.54	12.78	19.84
RMSEA-related fit indices										
RMSEA _{PS_B}	Mean	0.043	0.027	0.008	0.027	0.010	0.005	0.010	0.005	0.003
	SD	0.061	0.046	0.012	0.035	0.012	0.007	0.013	0.007	0.004
RMSEA _{PS_W}	Mean	0.048	0.024	0.007	0.027	0.008	0.004	0.010	0.005	0.003
	SD	0.093	0.052	0.009	0.045	0.011	0.006	0.013	0.007	0.004
RMSEA _{T_S_COV}	Mean	0.173	0.098	0.061	0.102	0.081	0.059	0.089	0.078	0.092
	SD	0.313	0.061	0.010	0.032	0.014	0.007	0.015	0.009	0.006
RMSEA _{T_S_Mean}	Mean	0.071	0.044	0.029	0.042	0.035	0.028	0.037	0.034	0.036
	SD	0.046	0.029	0.009	0.022	0.011	0.005	0.012	0.007	0.003
CFI-related fit indices										
CFI _{PS_B}	Mean	0.996	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	SD	0.025	0.008	0.000	0.005	0.000	0.000	0.000	0.000	0.000
CFI _{PS_W}	Mean	0.995	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	SD	0.046	0.016	0.000	0.010	0.000	0.000	0.000	0.000	0.000
CFI _{T_S_COV}	Mean	0.967	0.990	0.996	0.990	0.994	0.997	0.993	0.994	0.990
	SD	0.073	0.033	0.001	0.006	0.002	0.001	0.003	0.001	0.001
CFI _{T_S_Mean}	Mean	0.995	0.998	0.999	0.998	0.999	0.999	0.999	0.999	0.999
	SD	0.009	0.008	0.000	0.002	0.001	0.000	0.001	0.000	0.000
TLI-related fit indices										
TLI _{PS_B}	Mean	0.990	0.994	1.000	0.996	1.000	1.000	1.000	1.000	1.000

TLI _{PS_W}	SD	0.071	0.042	0.002	0.027	0.001	0.000	0.001	1.000	0.000
	Mean	0.979	0.994	1.000	0.994	1.000	1.000	1.000	1.000	1.000
TLI _{T_S_COV}	SD	0.281	0.079	0.001	0.051	0.001	0.000	0.001	0.001	0.000
	Mean	0.786	0.970	0.990	0.973	0.982	0.990	0.979	0.983	0.978
TLI _{T_S_Mean}	SD	4.211	0.147	0.003	0.018	0.006	0.002	0.007	0.004	0.002
	Mean	0.986	0.994	0.997	0.995	0.996	0.998	0.996	0.997	0.997
	SD	0.022	0.022	0.001	0.006	0.002	0.001	0.002	0.001	0.001
SRMR-related fit indices										
SRMR _B	Mean	0.014	0.008	0.005	0.008	0.005	0.004	0.005	0.003	0.003
	SD	0.010	0.005	0.003	0.005	0.003	0.002	0.003	0.002	0.001
SRMR _W	Mean	0.005	0.002	0.001	0.003	0.002	0.001	0.002	0.001	0.001
	SD	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
SRMR _{T_S_COV}	Mean	0.319	0.183	0.193	0.170	0.184	0.193	0.167	0.184	0.315
	SD	0.051	0.027	0.019	0.029	0.019	0.013	0.021	0.013	0.007
SRMR _{T_S_Mean}	Mean	0.062	0.070	0.072	0.066	0.070	0.071	0.069	0.072	0.068
	SD	0.034	0.030	0.021	0.028	0.021	0.015	0.021	0.015	0.008

∞ RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker–Lewis index. SRMR = standardized root mean square residual. Subscripted PS = partially saturated model method. Subscripted TS = target-specific fit indices. Subscripted B = between-level model. Subscripted W = within-level model. Subscripted COV = fit index for evaluating between-covariance structure. Subscripted MEAN = fit index for evaluating between-mean structure.

RMSEA-related fit indices: Linear Trajectory

In Table 4.3, means and standard deviations for the RMSEA-related fit indices, were calculated. When NG increased from 50 to 200 and unbalanced GS increased from 5/15 to 25/75, all means of $RMSEA_{PS_B}$ and $RMSEA_{PS_W}$ approached acceptable model fit (i.e., $<.06$) and standard deviation values decreased as sample size increased. Also, all $RMSEA_{T_S_COV}$ mean values did not approach levels indicative of acceptable model under the tested sample sizes. For the $RMSEA_{T_S_MEAN}$, values yielded an acceptable model, except under the smallest sample size condition (NG = 50, unbalanced GS = 5/15). As with other trajectories modeled, the standard deviation of $RMSEA_{T_S_COV}$ and $RMSEA_{T_S_MEAN}$ decreased as sample sizes increased.

CFI-related fit indices and TLI-related fit indices: Linear Trajectory

In Table 4.3, descriptive statistics for the CFI-related fit indices and TLI-related fit indices were provided. Means were indicative of good model fit (i.e., $>.95$) under almost all simulation conditions. Again, there was only one mean value, $TLI_{T_S_COV}$, that yielded a value under the stated cutoff (NG of 50, unbalanced GS of 5/15).

SRMR-related fit indices: Linear Trajectory

Table 4.3 reports means and standard deviations for the SRMR-related fit indices. As noted, all means of $SRMR_B$, $SRMR_W$, and $SRMR_{T_S_MEAN}$ produced values indicating acceptable model fit (i.e., $<.08$); however, means of $SRMR_{T_S_COV}$ were larger, approaching the cutoff of poor model fit under all simulation conditions. The standard deviations values of $SRMR_B$ and $SRMR_W$ were close to a value 0 and standard deviations of $SRMR_{T_S_COV}$ and $SRMR_{T_S_MEAN}$ were also small.

Analysis 2: ANOVA Review of Conditions

To determine factors that affected model fit indices, a three-way ANOVA with 3 (NG: 50, 100, and 200) x 3 (unbalanced GS: 5/15, 10/30, and 25/75) x 3 (accelerating, decelerating, and linear) levels was conducted, with each fit index as the outcome variable. For each factorial ANOVA, the eta-squared (η^2) was computed to provide the proportion of the variance accounted for by a particular design factor. Each simulation condition has the same number of simulated datasets, resulting in (balanced) orthogonal design factors. In this way, the Type III sum-of-squares from different factors were additive and non-overlapping, meaning the η^2 of each design factor could be calculated separately without considering other factors (Cohen et al, 2013). The Type III sum-of-squares for each fit index provided the variability of the fit index values across 1,000 replications under each simulation conditions. We computed eta-squared (η^2) by dividing the Type III sum-of-squares of a particular design factor by the Type III sum-of-squares of each fit index. η^2 provides the proportion of the variance accounted for by a particular design factor. Based on Cohen's (1988) suggestion, we adopted a medium η^2 of .0588 to denote an influential design factor on the fit indices' values (i.e., practical significance).

The η^2 for each design factor is presented in Table 4. We provide a visual representation of the influential design factors on the fit indices' values with boxplots in Figures 1 and 2, respectively. Box plots provide information on the variability and were useful when comparing distributions among conditions of a design factor to indicate how spread in the set of values, when a fit index had a η^2 larger than .0588.

Table 4.4 η^2 values from ANOVA design by fit index

Dependent Variables	Number of Group (NG)	Unbalanced Group Size (GS)	Trajectory
χ^2 Test Statistics			
$\chi^2_{PS_B}$	0.00	0.00	0.00
$\chi^2_{PS_W}$	0.00	0.00	0.00
$\chi^2_{T_S_COV}$	0.00	0.00	0.00
$\chi^2_{T_S_Mean}$	0.00	0.00	0.01
RMSEA-related fit indices			
RMSEA _{PS_B}	0.06	0.07	0.00
RMSEA _{PS_W}	0.05	0.06	0.00
RMSEA _{T_S_COV}	0.02	0.06	0.04
RMSEA _{T_S_Mean}	0.03	0.06	0.01
CFI-related fit indices			
CFI _{PS_B}	0.00	0.00	0.00
CFI _{PS_W}	0.00	0.00	0.00
CFI _{T_S_COV}	0.02	0.02	0.03
CFI _{T_S_Mean}	0.00	0.01	0.00
TLI-related fit indices			
TLI _{PS_B}	0.00	0.00	0.00
TLI _{PS_W}	0.00	0.00	0.00
TLI _{T_S_COV}	0.00	0.00	0.00
TLI _{T_S_Mean}	0.00	0.00	0.00
SRMR-related fit indices			
SRMR _B	0.13	0.14	0.02
SRMR _W	0.16	0.22	0.07
SRMR _{T_S_COV}	0.00	0.02	0.84
SRMR _{T_S_Mean}	0.00	0.01	0.01

RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker–Lewis index. SRMR = standardized root mean square residual. Subscripted PS = partially saturated model method. Subscripted TS = target-specific fit indices. Subscripted B = between-level model. Subscripted W = within-level model. Subscripted COV = fit index for evaluating between-covariance structure. Subscripted MEAN = fit index for evaluating between-mean structure. Highlighted (gray shaded cells) $\eta^2 \geq .05$

NG (50, 100, and 200)

Based on η^2 values in Table 4.4, only three indices: $RMSEA_{PS_B}$, $SRMR_B$, and $SRMR_W$, were impacted by NG, with η^2 values of .06, .13, and .26, respectively. Further, the boxplots in Figure 4.1 showed that the variabilities of $RMSEA_{PS_B}$ computed under all simulation conditions were large at lower sample sizes. As the NG increased, the median $RMSEA_{PS_B}$ decreased with values of 0.12, 0.04, to 0.001 associated with NG levels of 50, 100, and 200, respectively. The interquartile ranges also became smaller, indicating the values of $RMSEA_{PS_B}$ were less dispersed.

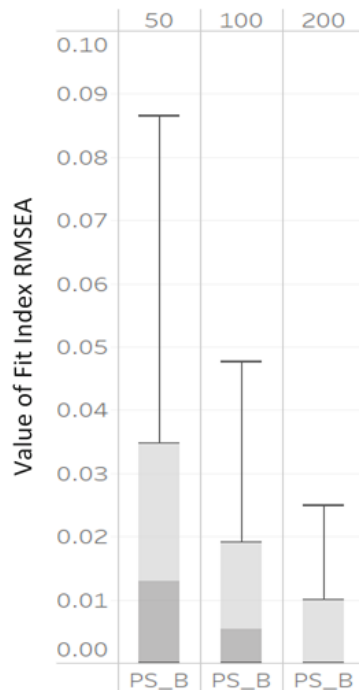


Figure 4.1 Box plot of $RMSEA_{PS_B}$ values derived from correctly specified MLGM models by NG (50, 100, and 200).

$SRMR_B$ and $SRMR_W$ also demonstrated large variability under the simulation conditions (shown in Figure 4.2). As the NG varied from 50, 100, to 200, the median

SRMR_B decreased from 0.007, to 0.005, to 0.0035 and the median SRMR_W decreased from 0.0025, to 0.0018, to 0.001. The interquartile ranges of all SRMR_B and SRMR_W also became smaller.

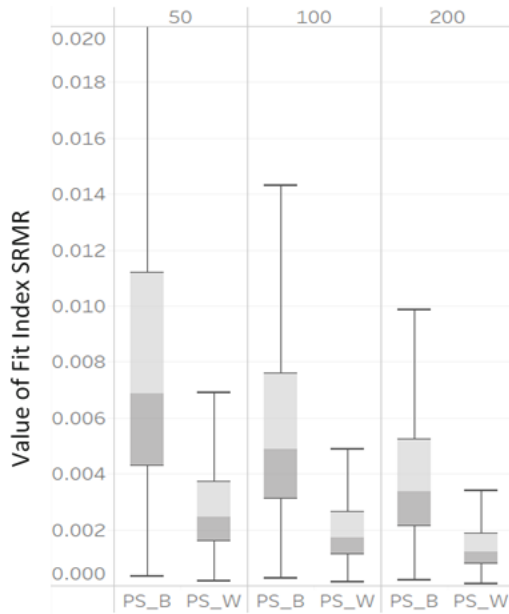


Figure 4.2 Box plot of SRMR_B and SRMR_W values derived from correctly specified MLGM models by NG (50, 100, and 200).

Unbalanced GS (5/15, 10/30, to 25/75)

From the ANOVA results presented in Table 4.4, η^2 indicated that RMSEA_{PS_B}, RMSEA_{PS_W}, RMSEA_{T_S_COV}, SRMR_B and SRMR_W were impacted by unbalanced GS ($\eta^2 = .07, .06, .06, .14$ and $.22$). There was no effect of unbalanced GS condition on all the other specific fit indices. As the unbalanced GS varied from 5/15, 10/30, to 25/75, the median RMSEA_{PS_B} would show poor fit at the smallest level (0.08) but not at later levels (0.04, 0.001 for 10/30 and 25/75, respectively). As the unbalanced GS increased, the median RMSEA_{PS_W} decreased with values of 0.08, 0.01, to 0.001 and the median

RMSEA_{T_S_COV} decreased with values of 0.12, 0.09, to 0.064 with unbalanced GS levels of 5/15, 10/30, and 25/75, respectively. Box plots are presented in Figure 4.3.

The interquartile ranges of all RMSEA_{PS_B}, RMSEA_{PS_W}, and RMSEA_{T_S_COV} also smaller, indicating the values of RMSEA_{PS_B}, RMSEA_{PS_W}, and RMSEA_{T_S_COV} were less dispersed. Both the η^2 and boxplots showed that RMSEA_{PS_B}, RMSEA_{PS_W}, and RMSEA_{T_S_COV} were affected by factor unbalanced GS and may indicate values indicating poor model fit for a correctly specified model fit.

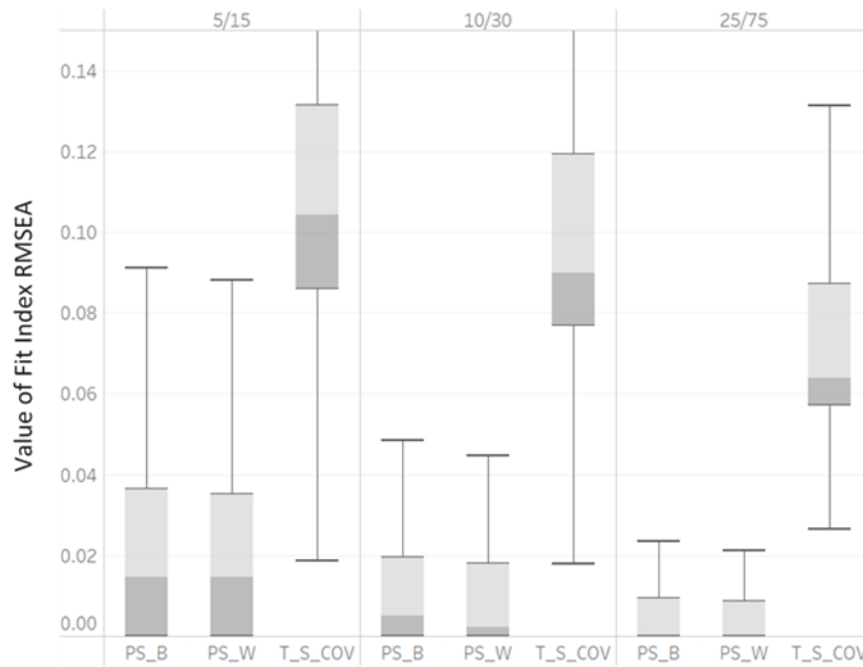


Figure 4.3 Box plot of RMSEA_{PS_B}, RMSEA_{PS_W}, and RMSEA_{T_S_COV} values derived from correctly specified MLGM models by unbalanced GS (5/15, 10/30, and 25/75).

Figure 4.4 demonstrated the variabilities of SRMR_B and SRMR_W computed under all simulation conditions were large. As the unbalanced GS increased, the median SRMR_B decreased (with values of 0.007, 0.005, to 0.0032) as did the median SRMR_W

decreased (values of 0.0024, 0.0018, to 0.001) for unbalanced GS levels of 5/15, 10/30, to 25/75, respectively.

The interquartile ranges of all $SRMR_B$ and $SRMR_W$ also became smaller, indicating the values of $SRMR_B$ and $SRMR_W$ were less dispersed. Both the η^2 and boxplots showed $SRMR_B$ and $SRMR_W$ were affected by factor unbalanced GS. With increase of the unbalanced GS, $SRMR_B$ and $SRMR_W$ showed lower model fit values.

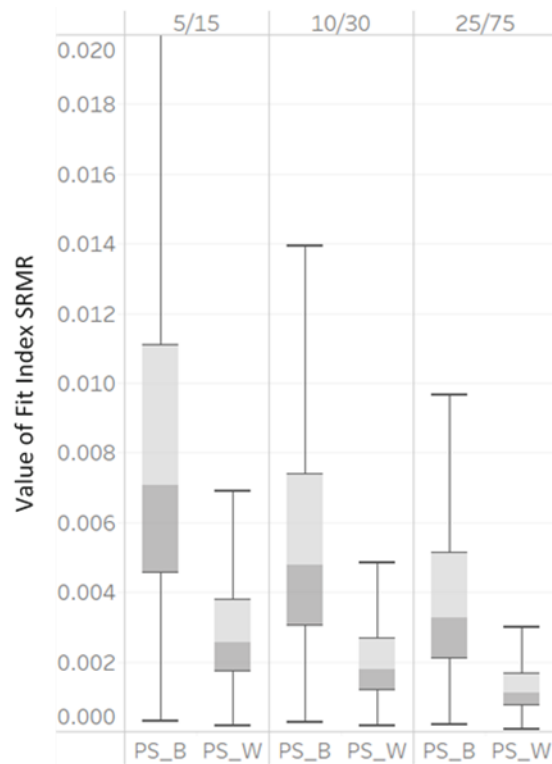


Figure 4.4 4 Box plot of $SRMR_B$ and $SRMR_W$ values derived from correctly specified MLGM models by unbalanced GS (5/15, 10/30, and 25/75).

Trajectory

$SRMR_W$ and $SRMR_{T_S_COV}$ also appeared to be strongly impacted by the type of change trajectory ($\eta^2 = .07$ and $.84$); however, there was no effect of trajectory condition

on all the other fit indices. The boxplots in Figures 4.5 and 4.6 showed that the distributions of $SRMR_W$ and $SRMR_{T_S_COV}$ were same for accelerating and decelerating and different from linear. As the trajectory varied across accelerating, decelerating, and linear change, the median $SRMR_W$ increased with values of 0.0015, 0.0015, to 0.0022; also, the median $SRMR_{T_S_COV}$ increased with values of 0.17, 0.17, to 0.31.

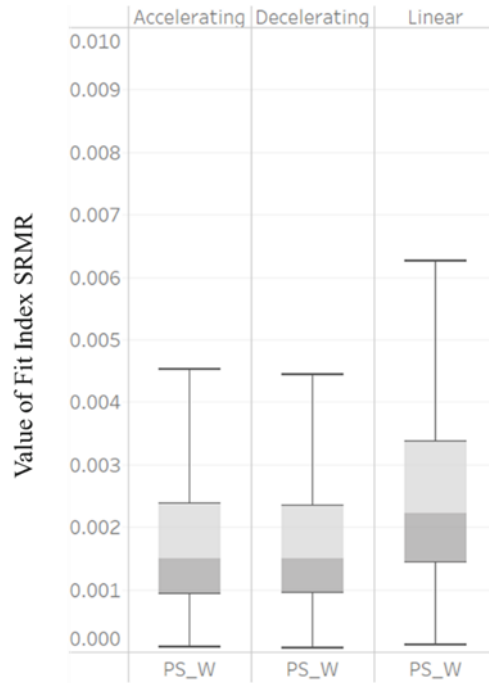


Figure 4.5 Box plot of $SRMR_W$ values derived from correctly specified MLGM models by trajectory (accelerating, decelerating, and linear)

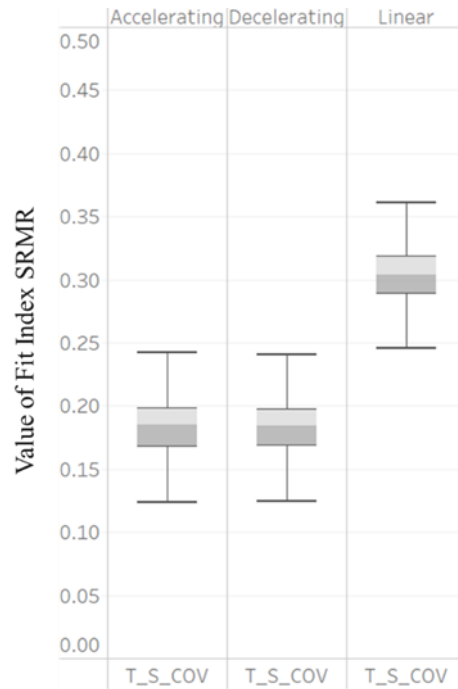


Figure 4.6 Box plot of $SRMR_{T_S_COV}$ values derived from correctly specified MLGM models by trajectory (accelerating, decelerating, and linear)

Summary

This chapter summarize the results of two analyses conducted. In the first set of analyses, the smallest sample size, $NC = 50$ and unbalanced $GS = 5$, were able to identify correctly specified between-level and within-level models. For χ^2 test statistics, mean values of $\chi^2_{PS_W}$ and $\chi^2_{PS_B}$ approached poor model fit when total sample size was small. A sample size over 1,000 was needed for $\chi^2_{PS_W}$ and $\chi^2_{PS_B}$ average values to appropriately identify correct between-level and within-level models. $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$ did not approach the values indicative of acceptable model fit when total sample size increased. All means of $RMSEA_{PS_B}$ and $RMSEA_{PS_W}$ approached values indicative of good model fit (i.e., $<.06$); however, $RMSEA_{T_S_COV}$ values did not show the different sample sizes. $RMSEA_{T_S_MEAN}$ values approached acceptable fit, except under a small sample size.

Means of the all CFI-related fit indices and TLI-related fit indices were indicative of good model fit (i.e., $>.95$) under all simulation conditions.

When NG was 50 and unbalanced GS was 5/15, only means of $TLI_{T_S_COV}$ illustrated less poor model fit, and this was apparent only for three cells of the design. All means of $SRMR_B$, $SRMR_W$, and $SRMR_{T_S_MEAN}$ produced values indicative of good model fit (i.e., $<.08$) across conditions. However, means of $SRMR_{T_S_COV}$ had approached values of poor model fit (i.e., $<.08$) under all simulation conditions.

In the second set of analyses, it was found that $RMSEA_{PS_B}$, $RMSEA_{PS_B}$, $RMSEA_{T_S_COV}$, $SRMR_B$, and $SRMR_W$ were influenced by different unbalanced GSs (5/15, 10/30, and 25/75). The other fit indices tested were not influenced by different unbalanced GS. As for different NGs (50, 100, and 200), fit indices $RMSEA_{PS_B}$, $SRMR_B$, and $SRMR_W$ were sensitive to this factor. The outcome of other fit indices did not influence by different NGs. In addition, it was found that modeling different types of change trajectories (accelerating, decelerating, and linear) influenced the values of $SRMR_W$ and $SRMR_{T_S_COV}$. Different trajectories did not influence the outcome of other fit indices.

CHAPTER 5

DISCUSSION

MLGM is rapidly gaining popularity in the analysis of nested data (e.g., students' data from same class) over time. With multilevel data, subjects in the same cluster usually have some degree of dependence. Ignoring the dependence of the data could lead to severe distortions of model fit and standard errors of estimates (Asparouhov & Muthén, 2016). MLGM extends the LGM model by accommodating the dependence between observations due to nested data.

When researchers are evaluating a MLGM, typical SEM model fit indices and commonly accepted cutoff values are used for interpretation. However, there are problems with using typical SEM model fit indices to judge MLGM fit (Hsu, Kwok, Acosta, & Lin, 2015). The typical SEM model fit indices are likely to be dominated by the model fit of the observation level model (e.g., within-level) because the sample size is typically much larger at the individual level than at the group level (e.g., between-level). As a solution to the problems encountered when using typical SEM model fit indices, researchers have developed level-specific and target-specific model fit indices to detect whether the poor fit of the hypothesized MLGM. Level-specific model fit indices separately evaluate different levels of MLGM, and target-specific model fit indices evaluate if poor model fit comes from the covariance or the mean structure of observation-level model.

Even though previous research has recommended level-specific and target-specific model fit indices to detect misspecification (i.e., a misspecified structure) at different levels or to detect misspecification at covariance and mean structures in observation level, Hsu (2019) only simulated a MLGM with equal number of subjects per cluster (balanced data) and modeled only one type of change (accelerating). However, unbalanced designs may be more common in empirical research. This refers to situations where the number of subjects per cluster is not equally distributed (Baumert & Lehmann, 2012). For example, the number of students per elementary school at different time points may be different, because students may move or transfer to other schools. Although the Hsu (2019)'s study indicated that the selected model fit indices worked well in his study, there has not yet been a comparison of multilevel model fit indices across different trajectories. As accelerating, decelerating, and linear change may be found in empirical research, comparison of model fit indices performance across different trajectories is needed.

To assist applied researchers, this study examined the performance of different level-specific and target-specific fit indices using a correctly specified MLGM. The purpose was to determine how these indices performed under different number of groups, unbalanced design, and trajectories. The level-specific and target-specific fit indices were calculated based on most commonly used fit statistics with SEM studies. Hsu (2019) has systematically examined the effectiveness of five level-specific fit indices ($\chi^2_{PS_B}$, $RMSEA_{PS_W}$, CFI_{PS_W} , TLI_{PS_W} , $SRMR_W$, $\chi^2_{PS_W}$, $RMSEA_{PS_B}$, CFI_{PS_B} , TLI_{PS_B} , and $SRMR_B$) and five target-specific fit indices ($\chi^2_{T_S_COV}$, $RMSEA_{T_S_COV}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, $SRMR_{T_S_COV}$, $\chi^2_{T_S_MEAN}$, $RMSEA_{T_S_Mean}$, $CFI_{T_S_Mean}$, $TLI_{T_S_Mean}$, and

SRMR_{T_S_Mean}). Typical cutoff criteria for SEM fit indices were employed to determine the NG and unbalanced GS needed to accurately identify the correctly specified model (i.e., RMSEA-related fit indices < .06; CFI- and TLI-related fit indices > .95; SRMR-related fit indices < .08; Boulton, 2011).

Discussion of Study 1: Descriptive Review of Conditions

As an extension of the previous balanced MLGM study (Hsu, 2019), we expected select fit indices to be less influenced by sampling errors arising from a small sample size, unbalanced design, and types of trajectories. In this first set of analysis, descriptive investigations were conducted to describe the basic features of the fit index, including mean and standard deviation. Results are discussed by fit index.

Chi-square. Generally, based on means and standard deviation values, $\chi^2_{PS_W}$ and $\chi^2_{PS_B}$ outperformed $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ under most sample size scenarios. The descriptive analyses showed that even though the hypothesized models were correctly specified, the average estimated $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ approached the expected value (i.e., model degrees of freedom) when NG = 100 and unbalanced GS = 10/30. $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ yielded over-rejection of correctly specified unbalanced MLGM when the sample size is small. In practice, given that researchers are not aware if the model is correctly specified, researchers are recommended to use $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ when the sample size is large enough (NG > 100 and unbalanced GS > 10/30) under one of the three trajectories examined here. For $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, average values were not indicative of acceptable model even at the largest sample size (NG = 200 and unbalanced GS = 25/75), meaning $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ did not approach the values indicative of acceptable model fit all the simulation conditions in our study. Based on our descriptive

findings, researchers need to be aware that using these $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ will cause the over-rejection of correctly specified unbalanced MLGM under situations similar to the sample sizes and trajectories used here. Based on these findings, $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ are not recommended to evaluate MLGM with unbalanced design. These findings are different from previous study under a balanced MLGM, where Hsu et al. (2019) showed that all χ^2 test statistics, $\chi^2_{PS_B}$, $\chi^2_{PS_W}$, $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, were detect correctly specified the misspecified MLGM with a balanced design and accelerating trajectory when NG is larger than 200 and GS is larger than 20. The behaviors of level-specific and target-specific χ^2 test statistics for MLGM is found to be different from what has been found with the performance of χ^2 , which should have mean equals to degree of freedom (df) and standard deviation equals to $2*df$. The reason for this difference may due to formulas for level-specific and target-specific χ^2 test statistics. Compared to the χ^2 formula, which evaluate the fitting of a whole model, the formulas for level-specific and target-specific χ^2 were based on the abstractions of the fittings of two models.

RMSEA. As for RMSEA-related fit indices, all means of $RMSEA_{PS_B}$ and $RMSEA_{PS_W}$ approached values indicative of good model fit (i.e., $<.06$) under the study conditions. $RMSEA_{PS_B}$ and $RMSEA_{PS_W}$ are recommended to researcher to evaluate MLGM with unbalanced design with similar conditions as to those used here. However, all means of $RMSEA_{T_S_COV}$ did not approach the values indicative of acceptable model under any tested simulation conditions. Based on our descriptive findings, researchers need to be aware that using $RMSEA_{T_S_COV}$ will cause the over-rejection of even a correctly specified unbalanced MLGM under all sample sizes and trajectories similar

with those studied. $RMSEA_{T_S_COV}$ is not recommended to evaluate MLGM with unbalanced design. For the $RMSEA_{T_S_MEAN}$, the data analysis showed that means of $RMSEA_{T_S_MEAN}$ would indicate an acceptable model, except the smallest sample size ($NG = 50$, unbalanced $GS = 5/15$). Researchers need to aware that $RMSEA_{T_S_MEAN}$ would result in the over-rejection of correctly specified unbalanced MLGM when the sample size is small. Given that researchers don't know if the tested model is correctly specified, researchers are recommended to use $RMSEA_{T_S_MEAN}$ when the sample size is large enough ($NG > 50$ and unbalanced $GS > 5/15$) under similar trajectories to ones used here.

CFI & TLI. Means of the CFI-related fit indices and TLI-related fit indices were indicative of good model fit (i.e., $>.95$) under almost all simulation conditions. Only one $TLI_{T_S_COV}$ yielded a value outside of this range (when NG was 50 and unbalanced GS was 5/15). Except $TLI_{T_S_COV}$, all the other CFI-related fit indices and TLI-related fit indices are recommended to evaluate MLGM with unbalanced design under all simulation conditions in our study. Researchers are recommended to use $TLI_{T_S_COV}$ when the sample size is large enough ($NG > 50$ and unbalanced $GS > 5/15$)

SRMR. For SRMR-related fit indices, all means of $SRMR_B$, $SRMR_w$, and $SRMR_{T_S_MEAN}$ demonstrated values illustrating good model-data fit (i.e., $<.08$). For researchers, $SRMR_B$, $SRMR_w$, and $SRMR_{T_S_MEAN}$ are recommended to evaluate MLGM with unbalanced design under all conditions similar to those used in our study. However, means of $SRMR_{T_S_COV}$ had approached values did not approach the values indicative of good model fit (i.e., $<.08$) under all simulation conditions.

Discussion of Study 2: ANOVA of Conditions

As an extension of previous work (e.g., Hsu, 2019), the second set of analyses evaluated the sensitivity of both level-specific and target-specific model fit indices to different design factors. We discussed the influence of different design factors by fit index.

NG. When considering the number of groups, results indicated that $RMSEA_{PS_B}$, $SRMR_B$ and $SRMR_W$ were influenced by NG (50, 100, 200). As the NG increased, the means of $RMSEA_{PS_B}$, $SRMR_B$ and $SRMR_W$ became smaller, showing that these three fit indices may not perform adequately for small NG. For this design factor, our results differed from previous level-specific and target-specific fit indices study conducted under balanced MLGM (Hsu, 2019). Hsu (2019) concluded that $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $RMSEA_{T_S_COV}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, $RMSEA_{T_S_Mean}$, and $TLI_{T_S_Mean}$ were significantly influenced by NG and shouldn't be used. However, only three fit indices were influenced by NG in our study ($RMSEA_{PS_B}$, $SRMR_B$ and $SRMR_W$).

A plausible explanation in the differences in which fit indices were affected could be due to the differences in GS. This study simulated the same NG conditions used with Hsu (2019)'s balanced MLGM: 50, 100, 200. However, our study simulated three unbalanced GS: 5/15, 10/30, and 25/75, where the averaged GS was 10, 20, and 50. The GS condition used in Hsu (2019)'s balanced design are: 5, 10, and 20. As shown in chapter 3, the formulas of fit indices are closely related to both NG and GS. With the increase of both NG and GS, the fit indices will indicate good model fit. As Hsu et al. (2019)'s research did not provide descriptive results, we suppose that $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $RMSEA_{T_S_COV}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, $RMSEA_{T_S_Mean}$, and $TLI_{T_S_Mean}$

may have indicated poor fit for a correctly specified MLGM when at the smallest sample sized (NG=50, GS=5).

Even though the $RMSEA_{PS_B}$, $SRMR_B$ and $SRMR_W$ affected by NG, the descriptive analysis showed that all these three fit indices show good model fit for all NS simulated in our study. In this way, we suppose that these fit indices may show a poor model fit for a correctly specified model with NG that is smaller than the smallest NG (50) simulated in our study. Based on the results, we recommend researchers to collect at least 50 for NG, if they want to use $RMSEA_{PS_B}$, $SRMR_B$ and $SRMR_W$. However, if the NG cannot be at least 50, $RMSEA_{PS_B}$, $SRMR_B$ and $SRMR_W$ are not recommended.

Although ANOVA results and evaluation of η^2 did not indicate that $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ were substantial influenced by NG (50, 100, 200), $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ may fail to identify correctly specified MLGM at the smallest NG simulated in our study (NG = 50), except at largest unbalanced GS used (25/75) and even with a moderate number of groups (NG = 100), $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ failed to identify correctly specified MLGM at the smallest unbalanced group size (unbalanced GS =5/15). Based on these results, $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ are not recommended with unbalanced MLGM designs when researchers have a NG smaller than 50. If researchers have moderate NG (around 100 cases), we recommend researchers to collect a GS larger than 5/15.

Unbalanced GS. ANOVA results and evaluation of η^2 indicated that $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $RMSEA_{T_S_COV}$, $SRMR_B$, and $SRMR_W$ were influenced by unbalanced GS (5/15, 10/30, and 25/75). As the unbalanced GS increased, the means of $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $SRMR_B$ and $SRMR_W$ became smaller. However, all means $RMSEA_{PS_W}$ indicated a poor model fit for a correctly specified MLGM. For unbalanced

GS design factor, our results differed from previous level-specific and target-specific fit indices results under a balanced MLGM (Hsu, 2019) as Hsu (2019) concluded CFI_{PS_B} , TLI_{PS_B} , $RMSEA_{PS_W}$, $CFI_{T_S_COV}$, $TLI_{T_S_COV}$, $RMSEA_{T_S_Mean}$, $CFI_{T_S_Mean}$, and $TLI_{T_S_Mean}$ were influenced by different GSs (5, 10, and 20).

Differences identified here may be due to modeling an unbalanced design. For example, as the unbalanced GS increases, the standard error of parameter estimates also become smaller in general (Hox, Maas, & Brinkhuis, 2010). Although an unbalanced multilevel longitudinal data is common with many educational research applications, as previous studies have not investigated the effects of this design on MLGM-based model fit indices. Hox et al (2010) simulated an unbalanced multilevel data at one time point to investigate the performance of a MCFA (multilevel confirmatory factor model) on unbalanced data. The results indicated that unbalanced data had little impact on the accuracy of model estimation of the within (i.e., individual) level model. However, for the between (group) level of multilevel confirmatory factor model, accuracy of model estimation tended to be underestimated. Different from unbalanced MCFA, unbalanced MLGM have different between-level and within-level at each time point and the unbalanced data have an influence on the model estimation of both levels.

Based on the above descriptive analysis, $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $SRMR_B$ and $SRMR_W$ still illustrate good model-data fit at the smallest unbalanced GS. However, these fit indices may show a poor model fit for a correctly specified model for unbalanced GS that is smaller than 5/15. If a researcher does not have an unbalanced group size at least of 5/15, $RMSEA_{PS_B}$, $RMSEA_{PS_W}$, $SRMR_B$ and $SRMR_W$ are not recommended. Further, $RMSEA_{T_S_COV}$ indicated poor model fit for a correctly specified MLGM and

was influenced by unbalanced GS, RMSEA_{T_S_COV} is not recommended to researchers under all conditions.

Even though ANOVA results and evaluation of η^2 did not indicate that $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ were substantially influenced by unbalanced GS. $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ may fail to identify correctly specified MLGM at the smallest unbalanced GS simulated in our study (unbalanced GS = 5/15) for all but the largest NG simulated (NG = 200). When group size is smaller, (NG = 50), $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ still fail to identify correctly specified MLGM at the moderate unbalanced GS in our study (unbalanced GS = 10/30). $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ are not recommended when researchers have small sample sizes. If researchers have a moderate unbalanced GS, such as around 10/30, we recommend researchers to collect a NG larger than 50.

Total sample size-combination of NG and unbalanced GS. As the combination of NG and unbalanced GS determine the total sample size in MLGMs, the total sample size may also influence the performance level-specific and target-specific model fit indices. Previous research illustrated that total sample size highly influenced the values of fit indices because of the issue of total sample size discrepancy, that is, the difference between a sample covariance matrix and the covariance matrix of the population (Fan & Sivo, 2005; Kaplan, 2009; Kline, 2011). When the total sample size is small, the discrepancy between the sample covariance matrix and the population covariance matrix will increase, and the discrepancy between a sample covariance matrix and covariance matrix reproduced by a correctly specified model will also increase. This large discrepancy will cause the fit indices to indicate correctly specified model as a poor fit model.

In the exploration of the performance of χ^2 test statistics, $\chi^2_{PS_B}$, $\chi^2_{PS_W}$, $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, we found that $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ statistics were not able to identify a correctly specified model, unless the total sample size was large (total sample size > 1,000). These findings were in line with Schermelleh-Engel et al.'s (2014) study, which studied the effectiveness of $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ when under total sample size conditions is large (total sample size = 6,000, 15,000, and 30,000). Schermelleh-Engel et al. (2014) found that $\chi^2_{PS_B}$ and $\chi^2_{PS_W}$ yielded Type I error rates lower than .05 and could definitively identify a correctly specified model when the total sample sizes at least 15,000, respectively. $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ would over-reject a correctly specified unbalanced MLGM at the largest total sample size in our study (total sample size > 15,000). As this total sample size is very large for typical empirical studies, $\chi^2_{T_S_COV}$ and $\chi^2_{T_S_MEAN}$ are not recommended to evaluate MLGM with unbalanced design under conditions similar to those studied here

As for RMSEA-related fit indices, with the increase of the total sample sizes, the means of RMSEA-related fit indices became small. The design used was a correctly specified MLGM, thus, the small values of RMSEA-related fit indices show that the fit indices can identify the model. These results are consistent with previous research. As was pointed out by Wu and West (2010), RMSEA-related fit indices vary systematically with the discrepancy between the sample matrix and the covariance matrix reproduced by a correctly specified model.

The previous balanced MLGM study also showed RMSEA-related fit indices for the between-model evaluation, including $RMSEA_{PS_B}$, $RMSEA_{T_S_COV}$, and $RMSEA_{T_S_MEAN}$, were likely to be affected by sampling errors (Hsu et al., 2019). As Hsu

et al. (2019)'s research did not show the results of descriptive analysis, we suppose that $RMSEA_{PS_B}$, $RMSEA_{T_S_COV}$, and $RMSEA_{T_S_MEAN}$ indicate poor fit for a correctly specified MLGM when total sample size is small (total sample size < 250).

The means of $SRMR_B$ and $SRMR_W$ became small with the increase of the total sample sizes. As we simulated a correctly specified MLGM, the small values of $SRMR_B$ and $SRMR_W$ showed good model fit and are recommended. $SRMR_{T_S_COV}$ would yield the over-rejection of correctly specified unbalanced MLGM at the largest total sample size (10000). As this total sample size is large, $SRMR_{T_S_COV}$ are not recommended for use with MLGM. $SRMR_{T_S_Mean}$ illustrated good fit under different total sample sizes simulated in our study; $SRMR_{T_S_Mean}$ are recommended for use.

Not all small total sample size raised evaluation concerns. We found that the CFI- and TFI-related fit indices were not affected by small total sample sizes and illustrated good model-data fit under all sample sizes simulated in our study. Only $TLI_{T_S_COV}$ yielded a poor model fit for a correctly specified model when total sample size is small (total sample size is 250). All the other CFI- and TFI-related fit indices behaved appropriately under different sample sizes simulated in our study when there is an unbalanced design. Except for smallest total sample size (total sample size is 250), all standard deviations were nearly 0, indicating the fit indices outputted in mostly all replications within each simulation conditions were able to identify the correctly specified model. The results are consistent with balanced MLGM. Hsu et al. (2019)'s balanced MLGM studies also showed that the most CFI- related fit indices were also not affected by different sample sizes. Hsu et al. (2019) also indicated that only two TLI-

related fit indices, TLI_{PS_B} and $TLI_{T_S_MEAN}$, needed a moderate total sample size (total sample size > 1000) under balanced MLGM.

Besides total sample size, the combination of NG and unbalanced GS might also influence the performance level-specific and target-specific model fit indices in other ways. When the NG is small, the amount of the sampling errors presents with the between-level related specific model fit indices will increase due to that small samples might commit a Type II error for $\chi^2_{PS_B}$. Based on formulas in chapter 3, the between-level related model fit indices are calculated based on $\chi^2_{PS_B}$. The large sampling error causes some between-level or target-specific fit indices for the between-covariance or between-mean structure to fail to identify a correctly specified between-level model. In this way, between-level related specific model fit indices, CFI_{PS_B} , TLI_{PS_B} , $\chi^2_{PS_B}$, $RMSEA_{PS_B}$, and $SRMR_B$, require NG at least to be large enough (e.g., $NG > 50$) to be able to identify correctly specific MLGM with unbalanced design. These results are in line with Hsu et al.'s (2019) balanced MLGM study. We also recommend applied researchers to consider NG when evaluating between-level related specific model fit indices, regardless of if the MLGM design is unbalanced or balanced. From the findings here and from previous research, applied researchers should implement a NG of at least 50, regardless of the GS and if the MLGM design is unbalanced or balanced.

In contrast to the between-model evaluation, both NG and unbalanced GS jointly determine the sample size of the within-level model and influence the performance of within-level specific model fit indices. As for the effect of unbalanced GS, $\chi^2_{PS_W}$ failed to identify correctly specified within-level model when the unbalanced GS was small (e.g., unbalanced $GS < 10/30$). When the unbalanced GS is small, the amount of the

sampling errors calculated in within-level specific model fit indices increases. Based on formulas noted in chapter 3, $\chi^2_{PS_W}$ need sufficiently sample size to yield accurate inferences. As for the effect of NG, $\chi^2_{PS_W}$ can fail to identify correctly specified within-level model when the NG is small (e.g., $NG < 50$). These findings were also consistent with Hsu et al.'s (2019) balanced MLGM study. Study findings were in line with Hsu et al.'s (2019) balanced MLGM study showed that all χ^2 test statistics, $\chi^2_{PS_B}$, $\chi^2_{PS_W}$, $\chi^2_{T_S_COV}$, and $\chi^2_{T_S_MEAN}$, were able to detect a misspecified model when NG is larger than 200 and GS is larger than 20. Based on the findings from our analysis and from previous research, we highly recommend applied researchers to collect average GS of 20 and consider NG when evaluating within-level related specific model fit indices, regardless of if the MLGM design is unbalanced or balanced.

Trajectory. ANOVA results and evaluation of η^2 indicated that $SRMR_W$ and $SRMR_{T_S_COV}$ were influenced by types of trajectories (accelerating, decelerating, and linear). For accelerating and decelerating, the means of $SRMR_W$ and $SRMR_{T_S_COV}$ were similar. However, compare to means of $SRMR_W$ and $SRMR_{T_S_COV}$ under accelerating and decelerating, the means of $SRMR_W$ and $SRMR_{T_S_COV}$ under linear increased. These results are new as previous Monte Carlo studies have only examined the MLGM with one type of trajectory (e.g., Hsu et al., 2019; Wu & West, 2010; Wu, West, & Taylor, 2009). We suppose that the reason for difference between the quadratic and linear MLGM results is due to the different matrix structures. Under an accelerating trajectory or a decelerating trajectory, the matrix structures of factor loadings for MLGM is a 5 x 3 matrix (i.e., a 5 timepoints x 3 growth parameters). However, for a linear trajectory, the matrix structures of factor loadings for MLGM (five time points) are (i.e., a 5 x 2 matrix),

where there are still five time points but only two latent growth parameters: latent intercept and linear slope. This is an assumption and we need future research to investigate this more.

As $SRMR_{T_S_COV}$ and $SRMR_W$ were influenced by types of trajectories, $SRMR_{T_S_COV}$ and $SRMR_W$ are not recommended to researchers under all conditions.

Overall Summary

Based on our simulation results and discussion earlier, when evaluate a between-level model of an unbalanced MLGM, we recommend applying CFI_{PS_B} and TLI_{PS_B} without any consideration. $\chi^2_{PS_B}$ need a total sample size larger than 1000 ($NG > 100$ and unbalanced $GS > 10/30$). $RMSEA_{PS_B}$ and $SRMR_B$ need a total sample equal or larger than the smallest simulated sample size in our study ($NG = 50$ and unbalanced $GS = 5/15$).

As for evaluating a within-level model of an unbalanced MLGM, we recommend applying CFI_{PS_W} and TLI_{PS_W} without any consideration. $\chi^2_{PS_W}$ need a total sample size larger than 1000 ($NG > 100$ and unbalanced $GS > 10/30$). $RMSEA_{PS_W}$ need a total sample equal or larger than the smallest simulated sample sizes in our study ($NG = 50$ and unbalanced $GS = 5/15$). As $SRMR_W$ were influenced by types of trajectories, $SRMR_W$ is not recommended to researchers under all conditions.

Based upon the results, $CFI_{T_S_COV}$ is recommended to evaluate a between-level covariance structure of unbalanced MLGM. $TLI_{T_S_COV}$ is recommended except the smallest sample sizes simulated in our study ($NG = 50$ and unbalanced $GS = 5/15$). $RMSEA_{T_S_COV}$ was significantly influenced by unbalanced GS and did not approach values indicative of good model fit (i.e., < 0.06) under all simulation conditions. $\chi^2_{T_S_COV}$

and $SRMR_{T_S_COV}$ are also not recommended to evaluate MLGM with unbalanced design as they also did not approach values indicative of good model fit under all simulation conditions

When evaluating the between-mean structure of unbalanced MLGM, CFI_{T_S_MEAN}, TLI_{T_S_MEAN}, and $SRMR_{T_S_MEAN}$ are recommended for applied researchers. RMSEA_{T_S_MEAN} is recommended when the total sample size larger than the smallest simulated sample sizes in our study (NG = 50 and unbalanced GS = 5/15). $\chi^2_{T_S_Mean}$ is also not recommended to evaluate MLGM with unbalanced design.

Table 5.1 *Recommended specific model fit indices for different model or structure of MLGM*

	χ^2 -	RMSEA-	CFI-	TFI-	SRMR-
Between-level	$\chi^2_{PS_B}$ (NG > 100 and unbalanced GS > 10/30)	RMSEA _{PS_B} (NG >50, unbalanced GS > 5/15)	CFI _{PS_B}	TLI _{PS_B}	SRMR _B (NG >50, unbalanced GS > 5/15)
Within-level	$\chi^2_{PS_W}$ (NG > 100 and unbalanced GS > 10/30)	RMSEA _{PS_B} (NG >50, unbalanced GS > 5/15)	CFI _{PS_W}	TLI _{PS_W}	SRMR _W (NG >50, unbalanced GS > 5/15)
Between-level covariance			CFI _{T_S_COV}	TLI _{T_S_COV} (NG >50, unbalanced GS > 5/15)	
Between-level mean		RMSEA _{T_S_MEAN} (NG >50, unbalanced GS > 5/15)	CFI _{T_S_MEAN}	TLI _{T_S_MEAN}	SRMR _{T_S_MEAN}

Limitations and Future Research

As it is impossible to consider all plausible situations in a single simulation study, generalizations beyond conditions simulated in our study should be made carefully. First, we conducted an MLGM with five-time points as the population model to compare to prior studies that also employ similar MLGMs. Future studies are needed to determine whether our findings can be generalized to other models, such as a model with different numbers of time points or with different types of trajectories (e.g., a piecewise-linear trajectory). In addition, necessary sample sizes for model fit indices under different population models (e.g., piecewise-linear trajectory models) can be investigated with future studies.

Second, in this study, misspecifications for the between and within models were not modeled. As there is not yet literature informing researchers about the performance of different level-specific and target-specific model fit indices in MLGM using unbalanced design and different trajectories, a correctly specified MLGM was simulated to fill this gap as a first step. As misspecifications in MLGM are possible, this aspect deserves systematic investigation in future simulation studies. By investigating different misspecifications, we can study the indices' sensitivity, which measures the extent to which specific fit indices could reflect the discrepancy between correctly specified models and misspecified hypothesized models. We expected desirable fit indices to demonstrate reasonable sensitivity to minor misspecifications and to be able to detect moderate misspecifications at both levels. Third, in our study, we considered a few numbers of design factors. Other design factors, such as a different number of time points and ICCs of repeated measures, can be manipulated in future studies.

In conclusion, social science researchers are often interested in understanding how characteristics of individuals or entities change over time. Longitudinal studies describe the changing pattern of characteristics of interest, allowing investigation of questions such as of how change comes about, how much change occurs, how the change process might differ across observations, and the determinants of that change over a set period. Therefore, this dissertation study filled previous research gaps by informing researchers about the performance of different level-specific and target-specific model fit indices in MLGM using unbalanced design and different trajectories. As the use of MLGMs is a relatively new field, it is hoped that this study helps further the field by informing researchers interested in using level-specific and target model fit indices for unbalanced design and with different growth patterns.

REFERENCES

- Asparouhov, T. (2011, September 20). Re: Simulation of multilevel SEM [Online forum comment]. Retrieved from <http://www.statmodel.com/discussion/messages/12/886.html?1538344873>
- Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Haring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163-192). Charlotte, NC: Information Age Publishing, Inc.
- Bandalos, D. L., & Leite, W. (2013). Use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 625–666). Charlotte, NC: Information Age.
- Baumert, J., Nagy, N. & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development*, 83(4), 1347-1367.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, 47, 563–592.

- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods and Research*, 16(1), 78-117.
- Boulton, A. (2011). *Fit index sensitivity in multilevel structural equation modeling* (Master's thesis, University of Kansas). Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/8047/Boulton_ku_0099M_11730_DATA_1.pdf?sequence=1&isAllowed=y
- Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models in social and behavioral research: Applications and data analysis methods. Newbury Park, CA: Sage.
- Carlson, E., Daley, T., Bitterman, A., Heinzen, H., Keller, B., Markowitz, J., & Riley, J. (2009). Early school transitions and the social behavior of children with disabilities: Selected finding from the pre-school elementary education longitudinal study wave 3 report. Retrieved from PEELS website: <https://www.peels.org/reports.asp>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. doi: 10.1080/10705510701301834
- Chen, L. J., Fox, K. R., Ku, P. W. & Wang, C. H. (2012). A longitudinal study of childhood obesity, weight status change, and subsequent academic performance in Taiwanese children. *Journal of School Health*, 82(9), 424-431.
- Chou, C. P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(3), 247-266.

- Chou, C. P., Yang, D., Pentz, M. A., & Hser, Y.-I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis*, 46(2), 213-225.
- Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education*, 29(1), 46-57.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Hillsdale: Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330–351.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38(4), 529-569.
- Dalton, S. & Schulz, H. (2006, April). The effects and treatment of missing data at two levels of hierarchical linear models. Paper presented at the Annual Meeting of American Educational Research Association. San Francisco, CA.
- DiStefano, C. (2016). Examining fit with structural equation models. In K. Schweizer & C. DiStefano (Eds.), *Psychological assessment – science and practice: Vol.3. Principles and methods of test construction: Standards and recent advances* (pp. 166-193). Boston, MA: Hogrefe Publishing.
- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23, 225-241.

- DiStefano, C., Mîndrilă, D., & Monrad, D. M. (2013). Investigating factorial invariance of teacher climate factors across school organizational levels. In M. S. Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 257-275). Rotterdam, NL: Sense Publishers.
- Donoghue, J. R., & Jenkins, F. (1992). A Monte Carlo study of the effects of model misspecification on HLM estimates (Report No. RR-92-69). Princeton, NJ: Educational Testing Service.
- Enders, C. K. (2001a). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1), 128-141.
- Enders, C. K. (2001b). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6(4), 352-370.
- Enders, C. K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 313-342). Greenwich, CT: Information Age Publishing.
- Enders, C. K. (2008). A note of the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling*, 15(3), 434-448.
- Enders, C.K. & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430-457.
- 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, 343–367.

- Gerbing, D. W., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods and Research*, 21, 132-160.
- Gibson, N. M. & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational and Psychological Measurement*, 63(2), 204-238.
- Goos, M., Damme, J. V., Onghena, P., Petry, K. & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology*, 51(3), 323-347.
- Graham, J. W., Hofer, S. M. & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value pattern: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2), 197-218.
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 80-100.
- Graham, J. W., & Coffman, D. L. (2012). Structural equation modeling with missing data. In Hoyle, R. H. (Eds.), *Handbook of Structural Equation Modeling* (pp. 277-294). New York, N.Y.: The Guilford Press.
- Grimm, K. J. (2008). Longitudinal association between reading and mathematics achievement, *Developmental Neuropsychology*, 33(3), 410-426.

- Harmsen, R., Helms-Lorenz, M., Maulana, R., & van Veen, K. (2019). The longitudinal effects of induction on beginning teachers' stress. *British journal of educational psychology*, 89(2), 259-287.
- Heck, R. H. & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, 46(3), 659-689.
- Hirvonen, R., Tolvanen, A., Aunola, K. & Nurmi, J. E. (2012). The developmental dynamics of task-avoidant behavior and math performance in kindergarten and elementary school. *Learning and Individual Differences*, 22(6), 715-723.
- Hox, J. J., Maas, C. J., & Brinkhuis, M. J. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica neerlandica*, 64(2), 157-170.
- Hsu, H. Y., Kwok, O., Acosta, S., & Lin, J. H. (2015). Detecting misspecified multilevel SEMs using common fit indices: A Monte Carlo study. *Multivariate Behavioral Research*, 50, 197–215. <https://doi.org/10.1080/00273171.2014.977429>
- Hsu, H. Y., Lin, J.H., Kwok, O. M., Acosta, S., & Willson, V. (2016). The impact of intraclass correlation on the effectiveness of level-specific fit indices in multilevel structural equation modeling: A Monte Carlo study. *Educational and Psychological Measurement*, 77, 5–31. <https://doi.org/10.1177/0013164416642823>
- Hsu, H. Y., Lin, J. J., Skidmore, S. T., & Kim, M. (2019). Evaluating fit indices in a multilevel latent growth curve model: A Monte Carlo study. *Behavior research methods*, 51(1), 172-194.

- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., ...
Leinwand, S. (2013). Education Longitudinal Study of 2002: Base-year to first follow-up data file documentation (NCES 2014-361). Washington, DC: National Center for Education Statistics.
- Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, 10, 128–141.
- Jamshidian, M., & Mata, M. (2008). Postmodeling sensitivity analysis to detect the effect of missing data mechanisms. *Multivariate Behavioral Research*, 43(3), 432-452.
- Jiang, H. (2014). *Missing data treatments in multilevel latent growth model: A monte carlo simulation study*. The Ohio State University.
- Jordan, N., Kaplan, D., Ramineni, C. & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45(3), 850-867.
- Jöreskog, K. G., and Sörbom, D. (1993). LISREL8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ: Erlbaum.
- Judge, S. & Watson, S. M. R. (2011). Longitudinal outcomes for mathematics achievement for students with learning disabilities. *The Journal of Educational Research*, 104(3), 147-157.
- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd). Thousand Oaks: Sage.

- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd). New York: Guilford Press.
- Kim, K. S. (1990). *Multilevel data analysis: A comparison of analytical alternatives*. Ph.D. dissertation, University of California, Los Angeles, United States -- California.
- Kromrey, J. D. & Hines, C. V. (1994). Nonrandomly missing data in multiple regression: An empirical comparison of common missing-data treatments. *Education and Psychological Measurement*, 54(3), 573-593.
- Kwon, H. (2011). *Monte Carlo Study of Missing Data Treatments for an Incomplete Level-2 Variable in Hierarchical Linear Models*. Ph.D. dissertation, The Ohio State University, United States -- Ohio.
- Lane, P. (2008). Handling drop-out in longitudinal clinical trial: A comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics*, 7(2), 93-106.
- Langkamp, D. L., Lehman, A., & Lemeshow, S., (2010). Techniques for handling missing data in secondary analyses of large surveys. *Academic Pediatrics*, 10(3), 205-210.
- Lee, J., Moon, S. & Hegar, R. L. (2010). Mathematics skills in early childhood: Exploring gender and ethnic patterns. *Child Indicators Research*, 4(3), 353-368.
- Leite, W. L., & Stapleton, L. M. (2011). Detecting growth shape misspecifications in latent growth models: An evaluation of fit indexes. *Journal of Experimental Education*, 79, 361-381.
- Lenkeit, J. (2012). Effectiveness measures for cross-sectional studies: A comparison of value-added models and contextualized attainment models, *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 24(1), 1-25.

- Li, J. (2010). Effects of Full Information Maximum Likelihood, Expectation Maximization, Multiple Imputation, and Similar Response Pattern Imputation on Structural Equation Modeling with Incomplete and Multivariate Nonnormal Data. Ph.D.dissertation, The Ohio State University, United States -- Ohio.
- Linda, N. Y., Lee, S. Y., & Poon, W. Y. (1993). Covariance structure analysis with three level data. *Computational Statistics & Data Analysis*, 15(2), 159-178.
- Liu, G. & Gould, A. L. (2002). Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *Journal of biopharmaceutical statistics*, 12(2), 207-266.
- Holder, C. (2015). *African american paternal figures' impact on their daughters' psychological adjustment and educational outcomes over time* (Doctoral dissertation, University of Alabama Libraries).
- Longford, N. T., & Muthén, B. (1992). Factor analysis for clustered observations. *Psychometrika*, 57(4), 581-597.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods*, 16, 444–467.
<https://doi.org/10.1037/a0024376>
- Ma, X., & Ma, L. (2004). Modeling stability of growth between mathematics and science achievement during middle and high school. *Evaluation Review*, 28, 104–122.
<https://doi.org/10.1177/0193841X03261025>
- Ma, X., & Wilkins, J. M. (2007). Mathematics coursework regulates growth in mathematics achievement. *Journal for Research in Mathematics Education*, 38, 230–257.

- Marsh, H. W., Hau, K. T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald. Multivariate applications book series* (pp. 275–340). Mahwah: Erlbaum.
- May, H. & Supovitz, J. A. (2006). Capturing the cumulative effects of school reform: An 11-year study of the impacts of America's choice on student achievement. *Educational Evaluation and Policy Analysis*, 28(3), 231-257.
- McArdle, J. & Epstein, D. (1987) Latent growth curves within developmental structural equation models. *Child Development*, 58(1), 110-133.
- Miller, J. D., Kimmel, L., Hoffer, T. B., & Nelson, C. (2000). *Longitudinal study of American youth: User's manual*. Evanston: Northwestern University, International Center for the Advancement of Scientific Literacy.
- Morgan, P. L., Farkas, G. & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, 44(5). 472-488.
- Moser, S. E., West, S. G. & Hughes, J. N. (2012). Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology*, 104(3), 603-621.
- Mundform, D. J., Schaffer, J., Kim, M.-J., Shaw, D., Thongteerparp. A. (2011). Number of replications required in Monte Carlo simulation studies: A synthesis of four studies. *Journal of Modern Applied Statistical Methods*, 10(1).
- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, 22(3), 376-398.
- Muthén, B. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (Ed.), *Latent Variable Modeling with Applications to Causality* (pp. 149-161). New York: Springer Verlag.

- Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park: Sage.
- Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *The handbook of advanced multilevel analysis*. New York: Routledge.
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267–316.
<https://doi.org/10.2307/271070>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th). Los Angeles: Muthén & Muthén.
- Nakai, M. (2011). Simulation study: Introduction of imputation methods for missing data in longitudinal analysis. *Applied mathematical science*, 5(57), 2807-2818.
- Neuman, W. L. (2009). *Social research methods: Qualitative and quantitative approaches* (7th). Boston: Allyn & Bacon.
- Olejnicki, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8 (4), 434-447.
- Olinsky, A., Shaw, C., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, 151(1), 53-79.
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, 19(1), 21-49.

- Peugh, J. L. & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestion for improvement. *Review of Educational Research*, 74 (4), 525-556.
- Poon, W. Y., & Lee, S. Y. (1992). Maximum likelihood and generalized least squares analyses of two-level structural equation models. *Statistics and Probability Letters*, 14(1), 25–30.
- Pornprasertmanit, S., Miller, P., Schoemann A., & Rosseel, Y. (2013). *semTools: Useful tools for structural equation modeling*. R package version 0.4-0.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage Publications.
- Prosser, R. (1991). *Multilevel Analysis with Incomplete Data: A Monte Carlo Comparison of Four Treatments*. Ph.D. dissertation, University of Toronto, Canada -- Toronto.
- Raykov, T., & Marcoulides, G. A. (2012). *A first course in structural equation modeling*. Routledge.
- Rescorla, L. & Rosenthal, A. S. (2004). Growth in standardized ability and achievement test scores from 3rd to 10th grade. *Journal of Educational Psychology*, 96(1), 85-96.
- Roberts, G., & Bryant, D. (2011). Early mathematics achievement trajectories: Englishlanguage learner and native English-speaker estimates, using the Early Childhood Longitudinal Survey.
- Rogosa, D., & Saner, H. (1995). Longitudinal data analysis examples with random coefficient models. *Journal of educational and behavioral statistics*, 20(2), 149-170.
- Roth, P. L. & Switzer, F. S. (1995). A Monte Carlo analysis of missing data techniques in a HRM setting. *Journal of Management*, 21(5), 1003-1023.

- Ryu, E. (2014). Model fit evaluation in multilevel structural equation models. *Frontiers in Psychology*, 5, 81:1–9. <https://doi.org/10.3389/fpsyg.2014.00081>
- Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, 16, 583 – 601. <https://doi.org/10.1080/10705510903203466>
- Savalei, V., & Bentler, P. M. (2005). A statistically justified pairwise ML method for incomplete nonnormal data: A comparison with direct ML and pairwise ADF. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(2), 183-214.
- Schafer, J. & Graham J. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7(2), 147-177.
- Schaffer, J. L. & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, 11(2), 437–457.
- Scheaffer, R. L., Mendenhall, W., III, & Ott, R. L. (2005). *Elementary survey sampling* (6th). Belmont: Thomson Brooks/Cole.
- Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology*, 5, 181:1–11. <https://doi.org/10.3389/fpsyg.2014.00181>
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling*. New York: Routledge.
- Shin, T., Davison, M., & Long, J. (2009). Effects of missing data methods in structural equation modeling with nonnormal longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(1), 70-98.

- Shin, T., Davison, M. L., Long, J. D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, 23(1), 92-100.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications Inc.
- Rappaport, L. M., Amstadter, A. B., & Neale, M. C. (2019). Model fit estimation for multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 0, 1-12. doi: 10.1080/10705511.2019.1620109
- Roberts, G., & Bryant, D. (2011). Early mathematics achievement trajectories: Englishlanguage learner and native English-speaker estimates, using the Early Childhood Longitudinal Survey.
- Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 583-601. doi: 10.1080/10705510903203466
- Savalei, V., & Bentler, P. M. (2005). A statistically justified pairwise ML method for incomplete nonnormal data: A comparison with direct ML and pairwise ADF. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(2), 183-214.
- Schafer (1997). *Analysis of incomplete multivariate data*, J.L. Chapman & Hall, London.
- Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology*, 5, 1-11. doi: 10.3389/fpsyg.2014.00181

- Schermelleh-Engel, K., Moosbrugger, H., & Müller, 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.
- Schumacker, R. E., & Lomax, R. G. (2010). A beginner's guide to structural equation modeling. New York: Routledge.
- Schweig, J. D. (2014). *Multilevel factor analysis by model segregation: Comparing the performance of maximum likelihood and robust test statistics* (Master's thesis, University of California). Retrieved from <https://escholarship.org/uc/item/34w830qf>
- Sessoms, J. C. L. (2019). *Level-specific fit index performance with diagonally weighted least squares estimation of multilevel structural equation models* (Doctoral dissertation, The University of North Carolina at Greensboro). Retrieved from https://libres.uncg.edu/ir/uncg/f/Sessoms_uncg_0154D_12703.pdf
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, 79(2), 310-334. doi: 10.1177/0013164418783530
- Shin, T., Davison, M. L., Long, J. D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, 23(1), 92-100.

- Simms, K. (2012). A hierarchical examination of the immigrant achievement gap: The additional exploratory power of nationality and educational selectivity over traditional explorations of race and socioeconomic status. *Journal of Advanced Academics*, 23(1), 72-98.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A Meta-analytic review of research 1990-2000. *Review of Educational Research*, 75(3), 417-453.
- Swoboda, C. C. & Kim, J. S. (2010, April). Multiple imputation methods with multilevel data: A simulation study. Paper presented at the Annual Meeting of American Educational Research Association, Denver, CO.
- Takahashi, Y. & Wisenbaker, J. (1999, April). A comparison of the effectiveness of missing data treatments in the context of structural equation modeling. Paper presented at the 1999 annual meeting of the American Educational Research Association, Montreal, Canada.
- Tanaka, J.S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, 58(1), 134-146.
- Twisk, J. & de Vente, W. (2002). Attrition in longitudinal studies: How to deal with missing data. *Journal of Clinical Epidemiology*, 55(4), 329-337.
- Voight, A., Shinn, M., & Nation, M. (2012). The longitudinal effects of residential mobility on the academic achievement of urban elementary and middle school students. *Educational Researcher*, 41(9), 385-392.

- Wang, Z., Rohrer, D., Chuang, C. C., Fujiki, M., Herman, K., & Reinke, W. (2015). Five methods to score the Teacher Observation of Classroom Adaptation Checklist and to examine group differences. *The Journal of Experimental Education*, 83(1), 24-50. doi: 10.1080/00220973.2013.876230
- Wayman, J. C. (2003, April). Multiple imputation for missing data: What is it and how can I use it? Paper presented at the 2003 annual meeting of the American Educational Research Association, Chicago, IL.
- Wen, X., Bulotsky-Shearer, R. J., Hahs-Vaughn, D. L., & Korfmacher, J. (2012). Head Start program quality: Examination of classroom quality and parent involvement in predicting children's vocabulary, literacy, and mathematics achievement trajectories. *Early Childhood Research Quarterly*, 27(4), 640-653.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209-231). New York, NY: Guilford.
- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology*, 8(1), 52-69.
- Wu, W., West, S. G., & Hughes, J. N. (2008). Effects of retention in first grade on children's achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology*, 100(4), 727-740.

- Wu, J. Y., & Kwok, O.m. (2012). Using SEM to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(1), 16-35. doi: 10.1080/10705511.2012.634703
- Wu, J. Y., Kwok, O., & Willson, V. L. (2015). Using design-based latent growth curve modeling with cluster-level predictor to address dependency. *Journal of Experimental Education*, 82, 431–454. <https://doi.org/10.1080/00220973.2013.876226>
- Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research*, 45, 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, 183–201. <https://doi.org/10.1037/a0015858>
- Youn, M. J., Leon, J., & Lee, K. J. (2011). The influence of maternal employment on children's learning growth and the role of parental involvement. *Early Child Development and Care*, 182(9), 1227-1246.
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30(1), 165–200.
- Zhang, Y. (2015). *Evaluating model fit for linear-linear piecewise multilevel latent growth curve models* (Doctoral dissertation, University of Maryland). Retrieved from <https://drum.lib.umd.edu/handle/1903/17259>

Zullig, K. J., Collins, R., Ghani, N., Patton, J. M., Huebner, E. S., & Ajamie, J. (2014).

Psychometric support of the school climate measure in a large, diverse sample of adolescents: A replication and extension. *Journal of School Health*, 84(2), 82-90

APPENDIX A

SELECTED SAS® CODES TO SIMULATE DATA

```
%macro datasimulation(n);
proc iml;
%do m = 1 %to &n*0.5;
seed = 5*&m+&NG*2*0.25*&k;
print seed;
data = j(&NG*&GS*2*0.25, 7, 1);
create a&m from data[colname={SchoolID StudentID T1 T2 T3 T4 T5}];
flag=0;
do j = 1 to &NG;
veci = j(&GS,1,1);
SchoolID = j*veci;
call randseed(seed);
m_g
=&&trajectory&k*(&g_b+RandNormal(1,{0,0,0},&p_b)`)+sqrt(&s_b)*rannor(j(5,1,seed
));
m_g12345 = veci*m_g`;
seed = seed + 1;
StudentID = 1000*SchoolID+(1:&GS)`;
call streaminit(seed);
T_12345 =m_g12345+(&& trajectory&k*&g_w)`+
RandNormal(&GS,{0,0,0},&p_w)*&&
trajectory&k`+rannor(j(&GS,5,seed))*sqrt(&s_w);
data =SchoolID||StudentID||T_12345;
append from data;
end;
%end;
close a&m;
%do m = &n*0.5 %to &n;
seed = 5*&p+&NG*2*0.75*&k;
print seed;
data = j(&NG*&GS*2*0.75, 7, 1);
create a&p from data[colname={SchoolID StudentID T1 T2 T3 T4 T5}];
flag=0;
do j = 1 to &NG;
veci = j(&GS,1,1);
SchoolID = j*veci;
call randseed(seed);
```

```

m_g=&&trajectory&k*(&g_b+RandNormal(1,{0,0,0},&p_b))+sqrt(&s_b)*rannor(j(5,1,
seed));
m_g12345 = veci*m_g`;
seed = seed + 1;
StudentID = 1000*SchoolID+(1:&GS)`;
call streaminit(seed);
T_12345 =m_g12345+(&& trajectory&k*&g_w)`+
RandNormal(&GS,{0,0,0},&p_w)*&&
trajectory&k`+rannor(j(&GS,5,seed))*sqrt(&s_w);
data =SchoolID||StudentID||T_12345;
append from data;
end;
%end;
close a&m;
data datafile.c&i;set
%do j =1 %to &NG;
    a&m
%end;;
Rep = &i;
run;
%end;
quit;
%mend datasimulation;
%datasimulation(5);
%macro datachange(n);
%let varlist=SchoolID StudentID T1 T2 T3 T4 T5;
%do i = 1 %to &n;
data _null_;
set datafile.a&i;
file "&textfile.\c&i..dat";
put &varlist;
run;
%end;
data datafile.datalist (drop=j);
put datanames $20.;
do j=1 to &n;
datanames = "c";
datanames = compress(datanames||j||".dat");
output;
end;
run;
%mend;

```

APPENDIX B

SAMPLE MPLUS DATA ANALYSIS CODE

```
TITLE: Monte Carlo two-level growth model
DATA: FILE IS replist.dat;
      TYPE = MONTECARLO;
VARIABLE: NAMES ARE SchoolID StudentID T1 T2 T3 T4 T5;
          USEVAR =T1 T2 T3 T4 T5;
          CLUSTER = SchoolID;

ANALYSIS: TYPE = TWOLEVEL;
MODEL: % WITHIN%
      T1 with T2 T3 T4 T5;
      T2 with T3 T4 T5;
      T3 with T4 T5;
      T4 with T5;
      T1;
      T2;
      T3;
      T4;
      T5;

      %BETWEEN%
      ib by T1@1 T2@1 T3@1 T4@1 T5@1;
      sb by T1@0 T2@1 T3@2 T4@3 T5@4;
      qb by T1@0 T2@-1 T3@-4 T4@-9 T5@-16;
      ib;
      sb;
      qb;
      ib with sb;
      ib with qb;
      sb with qb;
OUTPUT: svalues;
SAVEDATA: RESULTS ARE PS_B.dat;
```