LONGITUDINAL MEASUREMENT NON-INVARIANCE ON GROWTH PARAMETERS RECOVERY AND CLASSIFICATION ACCURACY IN GROWTH MIXTURE MODELING

by

Caiyan Zhang

B.S., Shanghai Ocean University, 2007

M.S., Rochester Institute of Technology, 2009

Submitted to the Graduate Faculty of

School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Research Methodology

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

School of Education

This dissertation was presented

by

Caiyan Zhang

It was defended on

November 19, 2015

and approved by

Suzanne Lane, Ph.D., Professor, Department of Psychology in Education

Clement A. Stone, Ph.D., Professor, Department of Psychology in Education

Aidan G.C. Wright, Ph.D., Assistant Professor, Department of Psychology

Dissertation Advisor: Feifei Ye, Ph.D., Assistant Professor, Department of Psychology in Education

LONGITUDINAL MEASUREMENT NON-INVARIANCE ON GROWTH PARAMETERS RECOVERY AND CLASSIFICATION ACCURACY IN GROWTH MIXTURE MODELING

Caiyan Zhang, PhD

University of Pittsburgh, 2015

First-order growth mixture model (1-GMM) has received increased attention over the past decade. It models class-specific latent growth trajectory and individual classification using composite scores computed over items of the same scale across multiple time points. By default, using composite scores assumes identical item-to-construct relationship over time (longitudinal measurement invariance; L-MI), which is not necessarily the case in research practice.

Violation of L-MI assumption has been studied using latent growth curve modeling where subjects are assumed to be sampled from one latent class. Deviation from L-MI assumption impacted the growth characteristics, thus producing invalid conclusions on the pattern of change. This study extends the prior research on the impact of L-MI violation to the situation where multiple latent classes exist. A Monte Carlo study was performed to examine how systematically varied measurement non-invariance impacted class-specific growth factor parameter recovery and classification accuracy. Five factors were systematically manipulated in studying the impact of L-MI assumption violation: directional change in non-invariant item intercepts, patterns of item loadings and item intercepts, percent of items containing a set of non-invariant item parameters, presence of time-adjacent within-item correlated measurement error, and latent class distances. Additionally, three GMMs were compared to assess their robustness against longitudinal measurement non-invariance, including 1-GMM, second order GMM with constrained measurement invariance, and second order GMM with freely estimated item factor loadings and item intercepts.

Accuracy, precision, Type I error, and power were examined on the slope factor parameter estimates. Additionally, mixture proportion and individual classification were assessed. Results show that the second order GMM with freely estimated item loadings and item intercepts was robust under various violation of L-MI and able to produce accurate estimates of slope factor parameters. Performance of the second order GMM with constrained measurement invariance on slope factor parameters recovery depended on the specific generating measurement non-invariance configuration. 1-GMM, on the other hand, was not able to recover the slope factor parameters with deviation from the L-MI assumption. With extremely unbalanced mixture proportions, class membership assignment was found not satisfactory regardless of simulated measurement non-invariance condition and analysis model.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS XVII

1.0 INTRODUCTION 1

1.1 INTRODUCTION OF GROWTH MIXTURE MODEL (GMM) 1

1.2 LONGITUDINAL MEASUREMENT INVARIANCE (L-MI) IN GMM 4

1.3 GOAL OF THE STUDY 7

1.4 RESEARCH QUESTIONS 8

1.5 SIGNIFICANCE OF THE STUDY 8

1.6 ORGANIZATION OF THE PAPER 9

2.0 LITERATURE REVIEW 11

2.1 GMM 11

2.1.1 First-order GMM, parameters, interpretation, assumptions and estimation 11

2.1.2 SOGMM, parameters, interpretation and its advantages over first-order GMM 16

2.2 LONGITUDINAL MEASUREMENT INVARIANCE (L-MI) 23

2.2.1 MI levels and their meanings in SOGMM 24

2.2.2 Partial L-MI 28

2.2.3 Identification and identification invariance 29

2.3 MEASUREMENT IN LATENT GROWTH 30

2.4 SUMMARY OF STUDIES IN L-MI IN LGM 32

2.5 CONCLUSION 39

3.0 METHODOLOGY 41

3.1 LITERATURE REVIEW OF GMM APPLICATION 41

3.2 FIXED FACTORS 43

3.3 MANIPULATED FACTORS 44

3.3.1 Contamination level of measurement non-invariance 44

3.3.2 Longitudinal measurement non-invariance pattern 45

3.3.3 Within-item correlated measurement error over time 48

3.3.4 Directional change on the non-invariant item intercepts 49

3.3.5 Latent class distance 49

3.3.6 Analysis model 50

3.3.7 Summary of the factors and their conditions 51

3.4 GENERATING MODELS 54

3.4.1 Item parameters 55

3.4.2 Unique factor variance 61

3.5 DATA GENERATION 61

3.6 DATA ANALYSIS 64

3.7 EVALUATION CRITERIA 65

3.8 DATA GENEARTION VALIDATION 68

4.0 RESULTS 70

4.1 CONVERGENCE 71

4.2 GROWTH CHARACTERISTICS RECOVERY UNDER MD = 5 72

4.2.1 Slope factor mean 73

4.2.1.1 Raw bias 73

4.2.1.2 RMSE 78

4.2.2 Slope factor variance 78

4.2.2.1 Relative bias 78

4.2.2.2 RMSE 81

4.2.3 Type I error and power 81

4.2.4 Classification 82

4.3 GROWTH CHARACTERISTICS RECOVERY UNDER MD = 1.5 84

4.3.1 Slope factor mean 85

4.3.1.1 Raw bias 85

4.3.1.2 RMSE 88

4.3.2 Slope factor variance 90

4.3.2.1 Relative bias 90

4.3.2.2 RMSE 90

4.3.3 Type I error and power 91

4.3.4 Classification 92

4.4 SUMMARY OF RESULTS 94

4.4.1 Analysis model 94

4.4.2 Correlated ME 94

4.4.3 Contamination level, L-MI pattern and non-invariant item intercepts directional change 95

4.4.4 Latent class distance 98

5.0 DISCUSSION 100

5.1 RQ1: IMPACT ON GROWTH FACTORS 101

5.2 RQ2: IMPACT ON CLASSIFICATION RATE 105

5.3 RQ3: BETWEEN-CLASS DIFFERENCE IN THE IMPACT ON GROWTH FACTORS 108

5.4 LIMITATIONS AND FUTURE RESEARCH 109

5.5 IMPLICATIONS AND RECOMMENDATIONS 110

APPENDIX A 112

BIBLIOGRAPHY 128

LIST OF TABLES

Table 1. Descriptive statistics for total sample size and sample size for the smallest class 43

Table 2. Design conditions for the study 52

Table 3. Acronyms for the study conditions to be used in presentation of results 53

Table 4. Population growth factor values for the two latent classes... 54

Table 5. Generating disturbance variance across conditions 55

Table 6. Population parameters of factor loadings for various L-MI patterns............................. 58

Table 7. Population parameters of item intercepts with increasing L-MI pattern 59

Table 8. Population parameters of item intercepts with decreasing L-MI pattern........................ 60

Table 9. Generating unique factor variance for conditions other than strict invariance............... 61

Table 10. Convergence rates by analysis models under each generating condition under MD=1.5. 72

Table 11. ANOVA results investigating class-specific growth rate estimates accuracy when MD = 5. 74

Table 12. RMSE on slope factor variance by each of the analysis model for each latent class. .. 81

Table 13. ANOVA results investigating class-specific growth parameter estimates accuracy when MD = 1.5. 86

Table 14. Relative bias in the slope factor variance by analysis model for each latent class under MD = 1.5. 90

Table 15. Raw count by analysis model for each latent class 93

Table 16. Summary table of recovery in growth characteristics by each analysis model when MD = 5. 96

Table 17. Summary table of recovery in growth characteristics by each analysis model when MD = 1.5 97

Table 18. Descriptive statistics of posterior probabilities for incorrect and correct classification by each analysis model 107

Table A1. Mean raw bias on slope factor mean for C1 under MD = 5 112

Table A2. Mean raw bias on slope factor mean for C2 under MD = 5 113

Table A3. Mean raw bias on slope factor mean for C1 under MD = 1.5 114

Table A4. Mean raw bias on slope factor mean for C2 under MD = 1.5 115

Table A5. Mean relative bias on slope factor variance for C1 under MD = 5........................... 116

Table A6. Mean relative bias on slope factor variance for C2 under MD = 5........................... 117

Table A7. Mean relative bias on slope factor variance for C1 under MD = 1.5........................ 118

Table A8. Mean relative bias on slope factor variance for C2 under MD = 1.5........................ 119

Table A9. Mean overall classification for C1 under MD = 5 120

Table A10. Mean overall classification for C2 under MD = 5 121

Table A11. Mean overall classification for C1 under MD = 1.5 122

Table A12. Mean overall classification for C2 under MD = 1.5 123

Table A13. Mean individual classification for C1 under MD = 5 124

Table A14. Mean individual classification for C2 under MD = 5 125

Table A15. Mean individual classification for C1 under MD = 1.5 126

Table A16. Mean individual classification for C2 under MD = 1.5 127

LIST OF FIGURES

Figure 1. A contrived example of the individual growth trajectories, the estimated class-specific average growth trajectories in LGM and GMM 6

Figure 2. PATH DIAGRAM FOR LINEAR GMM WITH P = 4, Q = 2, C = 2 16

Figure 3. Path diagram of 4 times occasions for a linear SOGMM with *j*=6, *r*=1, *s*=2 (item intercepts omitted) 22

Figure 4. Mean raw bias on class-specific slope factor mean under combinations of MI pattern and item intercepts change for each analysis model when MD = 5. Line of |.05| was drawn as reference. 75

Figure 5. RMSE on class-specific slope factor mean under combinations of MI pattern and item intercepts change for each analysis model when MD = 5. 79

Figure 6. Relative bias on slope factor variance under generating MI patterns for each analysis model when MD = 5. Line of |.05| was drawn as reference. 80

Figure 7. Empirical power rates detecting other-than-zero slope factor mean under varying conditions when MD = 5. ME = Measurement error correlation. Line of .80 was drawn as reference. 84

Figure 8. Mean raw bias on class-specific slope factor mean under combinations of MI pattern and item intercepts change for each analysis model when MD = 1.5. Line of |.05| was drawn as reference. 87

Figure 9. RMSE on slope factor mean under generating MI patterns for each analysis model when MD = 1.5. 89

Figure 10. RMSE on slope factor variance under generating MI patterns for each analysis model when MD = 1.5. 91

Figure 11. Empirical power rates detecting other-than-zero slope factor mean under varying conditions when MD = 1.5. 92

LIST OF EQUATIONS

EQUATION 1 11

EQUATION 2 12

EQUATION 3 12

EQUATION 4 12

EQUATION 5 12

EQUATION 6 12

EQUATION 7 12

EQUATION 8 13

EQUATION 9 13

EQUATION 10 13

EQUATION 11 13

EQUATION 12 17

EQUATION 13 18

EQUATION 14 18

EQUATION 15 18

EQUATION 16 18

EQUATION 17 18

EQUATION 18 19

EQUATION 19 ... 19
EQUATION 20 ... 20
EQUATION 21 ... 20
EQUATION 22 ... 20
EQUATION 23 ... 20
EQUATION 24 ... 20
EQUATION 25 ... 21
EQUATION 26 ... 21
EQUATION 27 ... 27
EQUATION 28 ... 27
EQUATION 29 ... 27
EQUATION 30 ... 30
EQUATION 31 ... 31
EQUATION 32 ... 31
EQUATION 33 ... 31
EQUATION 34 ... 31
EQUATION 35 ... 32
EQUATION 36 ... 32
EQUATION 37 ... 32
EQUATION 38 ... 55
EQUATION 39 ... 55
EQUATION 40 ... 62
EQUATION 41 ... 62

EQUATION 42 62

EQUATION 43 62

EQUATION 44 63

EQUATION 45 63

EQUATION 46 65

EQUATION 47 66

EQUATION 48 66

ACKNOWLEDGEMENTS

My sincere thanks to all my committee members including Dr. Feifei Ye, Dr. Suzanne Lane, Dr. Clement Stone and Dr. Aidan Wright, who have made the dissertation phase as easy as it can be. Particularly, to Dr. Feifei Ye, thank you for the intentional and unintentional pushes that I need and for having my back. My other thanks go to my supervisor at UPMC Sports Medicine, Dr. James Irrgang, and the Principal Investigators of the projects that I have worked on including Dr. Sabina Deitrick, Dr. Pat Chew, and Dr. Shannon Wanless. Thank you all for your kindness and being supportive for my effort in personal and professional growth.

I would like to thank my family who sustained me. To my dad, thank you for setting me out on this journey and the unselfish and generous support throughout the years. To my husband, thank you for believing in me and all the encouraging words when I felt down. To my son, Karl, thank you for all the hugs and kisses. To all of my friends inside and outside of the department, thank you for your unfading friendship and lasting company. It made this journey more fun and unforgettable. Lastly, in the dearest memory of Dr. Kevin Kim who guided me through the early stage of my research, my appreciation goes beyond words.

To all of you, without your presence and support, it would be impossible to come this far. You made me who I am. Thank you. I love all of you.

1.0 INTRODUCTION

1.1 INTRODUCTION OF GROWTH MIXTURE MODEL (GMM)

In various research areas, researchers have investigated heterogeneous qualitatively different developmental pathways. One area for example is in adolescent substance use, which has been reported in 2011 as the top public health problem in the United States according to The National Center on Addiction and Substance Abuse. The research area has received increased attention as misuse of substance in youth results in expensive consequences such as impaired health (e.g., substance use disorders, mental illness), increased risks of dangerous behaviors (e.g., unintended pregnancy, violence), and impaired academic and career performance (e.g., educational attainment, academic performance).

In studies of adolescent substance use development, the target is to describe how differently adolescent alcohol and drug use unfolds over time, and examine antecedents and consequences associated with different developmental courses. Growth mixture modeling (GMM; Muthén & Shedden, 1999) is well suited to these purposes. In general, GMM uses latent classes and latent growth factors to investigate different developmental growth trajectories (i.e., patterns of change) on an outcome variable among unobserved heterogeneous populations. It purposefully 1) identifies latent classes with distinct growth trajectories, 2) assesses class-specific growth trajectory shape by estimating the growth factor means, 3) estimates within-class variability around the growth factor means, 4) assigns latent class membership to individuals based on estimated latent class probabilities (i.e., posterior probabilities), 5) estimates mixture proportion (i.e., percentage of the individuals in one specific latent class), and 6) uncovers predictive covariates and distal outcome of latent class and/or latent growth factors (i.e., antecedence and consequence).

Researchers have taken advantage of GMM in modeling heterogeneous developmental substance use trajectories among adolescence. For example, Warner, White, and Johnson (2007) studied different developmental growth trajectories of problem drinking from adolescence into young adulthood. Participants' responses to the Revised Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 1989) were collected repeatedly to form the outcome with higher score indicating higher levels of alcohol related problems. Three latent classes, namely, adolescence-limit problem (ALP) class, no-or-low problem (NLP) class, and escalating problem (EP) class, were identified with each associated with a distinct nonlinear growth pattern. More specifically, the NLP class remained relatively low and stable over the course of the time while the ALP class was featured with an initial sharp increase followed by a gradual decline and then a sharp decline, reaching level as low as the NLP class by young adulthood. The EP class, on the other hand, though started at a lower level than the ALP class but consistently caught up over time, and eventually escalated and reached the highest level among the three classes at the end of the study period. The majority of the participants (66.2%) were found to pertain to the NLP class while the rest belonged to a problem drinker class (21.6% for ALP class, 12.1% for EP class).

As can be seen, meaningful clinical interpretations are closely related to the implication of extracting multiple developmental trajectories and classification of individuals into different latent classes. GMM enables examination in the change pattern for each adolescent subpopulation where at-risk youth with an increasing pattern of substance use can be detected. Moreover, the identified different patterns of onset and escalation in adolescent substance use is important as it might promote effective development and implementation of intervention and prevention programs (Colder, Campbell, Ruel, Richardson, & Flay, 2002; Hix-Small, Duncan, Duncan, & Okut, 2004).

Besides adolescence substance use, GMM has been widely applied to capturing specific developmental course, predictors and consequences in other research areas such as mental health (Broadbent et al., 2007; Dekker et al., 2007; Schaeffer, Petras, Ialongo, Poduska, & Kellam, 2003), and cognitive and language development (Landa, Gross, Stuart, & Bauman, 2012; Rescorla, Mirak, & Singh, 2000). As a modeling tool, GMM is fairly flexible and can be embedded in other modeling framework such as stage-sequential models (e.g.,Kim, 2012), mediation analysis (e.g., Lane, Bluestone, & Burke, 2013), and multilevel modeling (e.g.,Chen, Kwok, Willson, & Luo, 2010). Given the flexibility, GMM has received increased attention over the past decade. According to the PsycINFO database (retrieved on 10/25/2014), the number of publications with GMM application was eight-folded in 2013 as compared to 2004 (i.e., 57 in 2013 as opposed to 7 in 2004).

With the increased popularity in empirical research, methodological researchers have made an endeavor to expand and extend GMM's utilities such as 1) GMM with nonlinear growth patterns (e.g., Estabrook, Ram, & Grimm, 2010; Kohli, Harring, & Hancock, 2013; Zopluoglu, Harring, & Kohli, 2014), 2) robustness of GMM to non-ignorable missing data (e.g., Lu, 2011; Zhang, Lu, & Lubke, 2011), and 3) alternative Bayesian estimators of GMM (e.g., Depaoli, 2010, 2012, 2013; Depaoli & Boyajian, 2014). However, there is a lack of research in studying the effects of measurement model assumption violation in GMM, particularly, consequences from deviation of longitudinal measurement invariance (L-MI) assumption.

1.2 LONGITUDINAL MEASUREMENT INVARIANCE (L-MI) IN GMM

L-MI is a prerequisite for valid interpretation of pattern of change over time in GMM. L-MI is evolved from measurement invariance (MI), which is rooted in classical test theory (CTT; Lord & Novick, 1968). MI assumption is used to evaluate the observed (i.e., manifest) variables' measurement equivalence across groups. It has been researched extensively with the focus on measurement equivalence between independent groups in a cross-sectional framework, such as gender (e.g., Byrne, 1988; Byrne, Baron, & Balev, 1996) and race (e.g., Chan, 1997; Collins & Gleaves, 1998). The assumption implies the score on the instrument is independent of any variables other than the person's value on the construct (Eid & Diener, 2006). When MI assumption holds, it indicates that respondents from different groups conceptualize a given instrument in a similar way. When MI assumption is violated, the observed group difference is confounded with the differential function of the instrument between the groups. It would be unknown whether the observed group difference is merely a consequence of violation of the assumption or a true difference on the construct between the groups. Hence, with violation of MI between independent groups, the meaningful interpretation of between-group difference is in jeopardy (Bollen, 1989; Drasgow, 1984, 1987).

The logic generalizes to longitudinal studies where the interest is to examine the change in the construct(s). Items of the instrument repeatedly administrated need to have consistent meaning over time (i.e., meeting L-MI assumption) to ensure the manifested change on the instrument over time is only due to the true change on the underlying construct rather than the differential functioning of the items over time.

In GMM, latent classes of subjects are formed with different change patterns. By default, GMM assumes L-MI by using the score on the same instrument in modeling change patterns over time. However, the use of the same instrument does not guarantee the tenability of L-MI assumption. At the instrument level, the relation between the construct and the instrument might have a temporal change, calling for changed operational definition of the construct (e.g., Caspi & Roberts, 2001; Fergusson, Horwood, Caspi, Moffitt, & Silva, 1996; Pajer, 1998). At the item level, relationship between items and the construct does not always stay unchanged. For example, the same response category of one item might mean different things over time for one subject. Regardless at which level, as long as one set of items can be interpreted differently over time, the concluded change pattern in the underlying construct would be questionable. Hence, in GMM, the crux in making sound conclusion in patterns of change in the construct relies on the maintenance of L-MI assumption.

Impact of L-MI violation has been studied in several longitudinal models to study intra- and inter-individual change, such as latent growth models (Leite, 2007; Olivera-Aguilar, 2013; Wirth, 2008) and autoregressive quasi-simplex model (Olivera-Aguilar, 2013). Latent growth modeling (LGM; Meredith & Tisak, 1990) resembles GMM in using continuous latent growth factors to represent the average growth trajectory and the variability around it. However, LGM assumes all individuals are from one homogeneous population while GMM assumes individuals are sampled from different latent classes.

The heterogeneous population assumption in GMM implies at least two dissimilarities from LGM. First, there are multiple average growth trajectories, rather than one as in LGM, modeled in GMM with each associated with one latent class. Secondly, individuals in GMM can be classified based on their estimated posterior probabilities of the latent class membership. An example demonstrating these dissimilarities between LGM and GMM is graphed. Growth trajectory over 4 occasions is plotted for each individual (left panel of Figure 1). The individual growth trajectories illustrate within-individual changes over time. Meanwhile, the trajectories, collectively, indicate presence of between-individual differences in the changes on the outcome as some of the trajectories increase while the others remain stable over time.

GMM is able to capture the two general growth patterns by using two latent classes. With C1 denoting class 1 and C2 for class 2, class-specific average growth trajectory is graphed in the right panel. The individuals whose growth trajectories resemble a stable pattern would be classified as from C1 and the rest who take an increasing pattern would be classified as from C2. Had the population been assumed as homogeneous in LGM, one average growth trajectory (i.e., C0) would have been estimated and every individual would have been classified into C0. Hence, GMM not only extracts multiple average growth trajectories but also allows accurate classification of individuals into their appropriate sampled class.

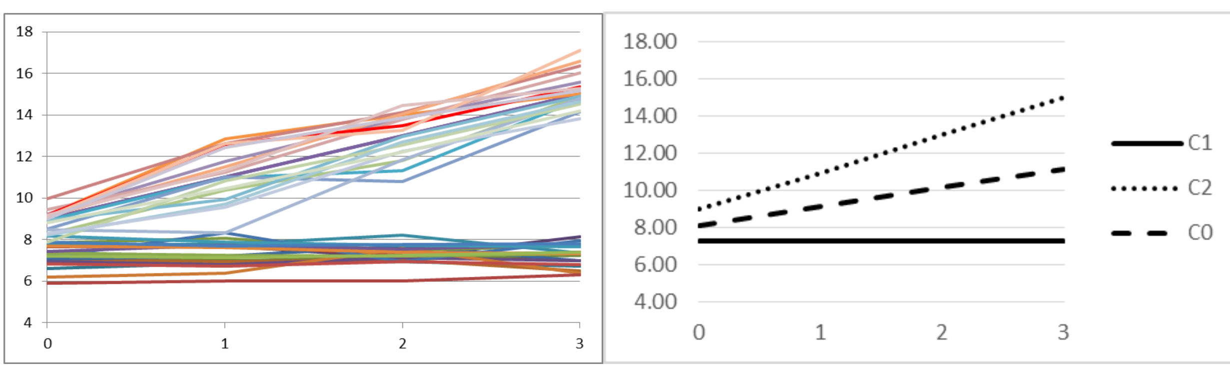

Figure 1. A contrived example of the individual growth trajectories, the estimated class-specific average growth trajectories in LGM and GMM

In LGM, violation of L-MI was found to alter growth characteristics including shape of the growth trajectory and growth factor estimates (Leite, 2007; Olivera-Aguilar, 2013; Wirth, 2008). With higher degree of violation such as more non-invariant items, a LGM model without capability of accommodating L-MI deviation produced biased growth factors estimates. With no item parameter maintaining L-MI assumption, none of the growth factors' parameters, including their means, variances, and covariance, were accurately or efficiently recovered (Wirth, 2008).

1.3 GOAL OF THE STUDY

As one average growth trajectory is subject to the influence of L-MI violation in LGM, we expect that deviation from L-MI will impact the growth characteristics recovery in GMM with multiple latent classes. Up to date, there has been no research looking into the effects of L-MI violation in GMM. Hence, this study aims at bridging the research gap by extending the prior studies on the L-MI violation in LGM to GMM. More specifically, the study will first assess the impact of L-MI violation in GMM on growth characteristics recovery including growth factor estimates and classification accuracy. Secondly, the study will evaluate the robustness of alternative GMM models under various deviations from L-MI assumption. While most of the applied studies with GMM use the first-order GMM model, it is impossible to assess L-MI assumption as composite scores instead of item scores are modeled. In contrast, a second-order GMM (SOGMM; Ram & Grimm, 2009) can test L-MI explicitly using item scores and simultaneously modeling growth trajectories for multiple latent classes. Consequently, a SOGMM was used to generate item scores with different levels of L-MI. Both first- and second-order models, assuming L-MI on all items, and a second-order model, assuming L-MI on one item, were applied to the simulated data to compare their performance under impact of longitudinal measurement non-invariance.

1.4 RESEARCH QUESTIONS

Three research questions are addressed in this study:

1) Are the concluded biased estimates of growth factors in LGM with longitudinal measurement non-invariance generalizable to GMM and if yes, what factors contribute to the bias?
2) Does the longitudinal measurement non-invariance impact the classification accuracy and if yes, what factors contribute to the classification inaccuracy?
3) Do the factors affect the growth factors' estimation for each latent class in the same way with the same magnitude?

1.5 SIGNIFICANCE OF THE STUDY

One of the primary uses of GMM is to identify growth characteristics such as average growth rate for multiple subpopulations. Classes associated with particular growth trend and the individuals within these classes are often identified, for example, in adolescent substance use developmental studies for intervention and prevention purposes. L-MI is a prerequisite for measuring change over time in GMM. Longitudinal measurement non-invariance has been shown to affect the performance of LGM, and thus is projected to impact the performance of

GMM. There has been no research on examining the impact of violation of L-MI on growth factor recovery and classification rate in GMM. This study has significant implications for both empirical users of the model and methodological researchers. Measurement non-invariance is prevalent in applied research. It is important to understand how the growth factors and the classification rates are affected by different configurations of L-MI such as number/percentage of non-invariant items and combinations of non-invariant item parameters. This will provide some guidance in the degree of confidence in interpreting primary questions that GMM addresses under varying configurations of longitudinal measurement non-invariance. Additionally, the comparison of the frequently adopted first-order GMM and SOGMMs will inform applied researchers to choose the optimal model in presence of longitudinal measurement non-invariance.

1.6 ORGANIZATION OF THE PAPER

To proceed, the second chapter starts with an introduction to first-order GMM including its model parametrization, parameter interpretation, assumptions, and estimation, followed by SOGMM with comparable sections and a conclusion on SOGMM's similarities and advantages to the first-order GMM. L-MI will then be introduced in the framework of SOGMM. It is followed by the description of partial L-MI, identification, identification invariance, and how longitudinal measurement non-invariance might affect latent growth factors parameters using examples. Studies examining unaccounted L-MI in LGM will be summarized regarding their strengths, weakness. Chapter 3 describes the Monte Carlo study including the research design, data generation, data analysis, data validation and the evaluation criteria. Chapter 4 presents results from the Monte Carlo study. Finally, Chapter 5 summarizes the findings with discussion of the results, limitations, implication in applied research and future research.

2.0 LITERATURE REVIEW

This study investigates the comparative performance of SOGMM and 1st-order GMM in recovery of growth parameter estimates and classification accuracy with varying degrees and L-MI configurations. This chapter discusses 1st-order GMM, SOGMM, longitudinal ME/I in SOGMM, in details, followed by examples and literature review in modeling latent growth when there is measurement non-invariance in LGM.

2.1 GMM

2.1.1 First-order GMM, parameters, interpretation, assumptions and estimation

GMM is used to identify latent classes with heterogeneous growth patterns. Consider the standard unconditional linear GMM on an outcome y_{ti} of individual *i* at time *t*:

$$y_{ti} = \lambda\eta_{0i}^{(c)} + \lambda_t\eta_{1i}^{(c)} + \varepsilon_{ti}\,,\ \varepsilon_{ti} \sim N(0,\theta)\,, \tag{1}$$

where the outcome (y_{ti}) is defined as the influence of two continuous latent variables (or latent growth factors η_{0i} and η_{1i}) through a function of time (factor loadings of λ and λ_t) and residuals on the observed outcome (ε_{ti}). Factor loadings of λ are fixed at 1 across time and λ_t are set at 0, 1, 2, …, *t*-1 allowing for interpretation of η_0 as the latent intercept factor at the initial time point and

η_1 as the rate of linear change over time (i.e., slope factor). Both latent intercept and linear slope are treated as random effects so that growth trajectories at latent class level and individual level can be modeled. More specifically, the latent intercept and linear slope factors, respectively, are expressed as a composite of growth mean parameters (α_0 and α_1) and individual deviations (ς_{0i} and ς_{1i}) away from these means:

$$\eta_{0i}^{(c)} = \alpha_0^{(c)} + \varsigma_{0i}, \tag{2}$$

$$\eta_{1i}^{(c)} = \alpha_1^{(c)} + \varsigma_{1i}, \tag{3}$$

$$\text{with } \begin{bmatrix} \varsigma_{0i} \\ \varsigma_{1i} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{\eta_0} & \omega_{\eta_0\eta_1} \\ & \omega_{\eta_1} \end{pmatrix}\right], \tag{4}$$

where c in the parentheses indicates the particular parameter is for latent class c. Growth mean parameters (i.e., mean intercept and mean linear slope), modeled as fixed effects, jointly define the average growth trajectory pooling of individuals within each latent class. The individual deviations from the respective means, on the other hand, are treated as random, enabling forming of individual growth trajectories so that each individual has his/her own intercept and linear slope.

The equations (1-4) in matrix notation can be written as:

$$\mathbf{y_i} = \mathbf{\Lambda}\boldsymbol{\eta}_\mathbf{i}^{(\mathbf{c})} + \boldsymbol{\varepsilon}_\mathbf{i}, \tag{5}$$

$$\text{with } \boldsymbol{\eta}_\mathbf{i}^{(\mathbf{c})} \sim \mathbf{N}\left(\boldsymbol{\alpha}^{(\mathbf{c})}, \mathbf{\Psi}\right), \tag{6}$$

$$\text{and } \boldsymbol{\varepsilon}_\mathbf{i} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Theta}), \tag{7}$$

where $\mathbf{y_i}$ is a $p \times 1$ vector of repeated measured outcome for individual i, p is the number of measurement, $\boldsymbol{\eta}$ is a $q \times 1$ vector of class-specific latent growth factors and q is the number of latent growth factors. The factor loading matrix $\mathbf{\Lambda}$ has a dimension of $p \times q$, and $\boldsymbol{\varepsilon}$ is a

$p \times 1$ vector of residuals on the outcome. Class-specific growth mean parameters are contained in $\boldsymbol{\alpha}$ which is a $q \times 1$ vector. Variance-covariance matrix of the latent growth factors $\boldsymbol{\Psi}$ has a dimension of $q \times q$ and the variance-covariance matrix of residuals on the outcome, $\boldsymbol{\Theta}$, has a dimension of $p \times p$ and is a diagonal matrix (i.e., off-diagonal elements are constrained to be 0).

Estimation of the above parameters is obtained through finite mixture model framework. Given that the observed outcome $\mathbf{y_i}$ is defined as the linear combination of normally distributed growth factors and residuals on the outcome, $\mathbf{y_i}$ is then distributed multivariate normally with the probability density function as:

$$f(\mathbf{y_i}) = \sum_{c=1}^{C} \pi^{(c)} \phi^{(c)} [\mathbf{y_i}; \boldsymbol{\mu}^{(\mathbf{c})}(\boldsymbol{\theta}^{(\mathbf{c})}), \boldsymbol{\Sigma}(\boldsymbol{\theta}^{(\mathbf{c})})], \tag{8}$$

$$\text{with } \sum_{c=1}^{C} \pi^{(c)} = 1, \tag{9}$$

where $\pi^{(c)}$ is the mixture proportion (or the average probability that an individual is drawn from latent class *c*). $\phi^{(c)}$ is the class-specific multivariate normal probability density function for $\mathbf{y_i}$ with model-implied mean vector and covariance matrix as:

$$\boldsymbol{\mu}^{(\mathbf{c})}(\boldsymbol{\theta}^{(\mathbf{c})}) = \boldsymbol{\Lambda}\boldsymbol{\alpha}^{(\mathbf{c})}, \tag{10}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}^{(\mathbf{c})}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{11}$$

where $\boldsymbol{\theta}$ is a vector containing all the model parameters to be estimated. The covariance matrix is class invariant (i.e., without superscript *c*) since the factor loadings ($\boldsymbol{\Lambda}$), covariance matrices of the growth factors ($\boldsymbol{\Psi}$), and residuals on the outcome ($\boldsymbol{\Theta}$) are assumed to be the same between the classes. Though a more general model could have been defined[1], this paper considers only the difference in the class-specific growth mean parameters.

Maximum likelihood estimator (MLE) through expectation-maximum algorithm (EM; Dempster, Laird, & Rubin, 1977) is the most frequently used estimation method to obtain the parameters contained in $\boldsymbol{\theta}$ (McLachlan & Peel, 2000). Since the latent class membership (c) is not directly observable, the main idea is to augment the data with a vector of binary variables indicating latent class membership so that the likelihood function can be constructed similarly as regular likelihood function with complete data. An estimate of the probability that an individual belongs to a class given the observed data (i.e., posterior probability) is used to calculate the expectations for class membership in the Expectation-step (i.e., E-step). In the Maximization-step (i.e., M-step), the posterior probability from the E-step is used to maximize multiple expected complete-data log likelihood functions. The E- and M-step then keep alternating between each other with the input from the E-step to produce the parameters in the M-step which is used for the input for the next E-step, until convergence (the difference is negligible between the log likelihood functions evaluated respectively with parameters at an iteration and its proceeding iteration). Then MLE ($\hat{\theta}$) is found that maximizes the likelihood in which the observed data would have been drawn from the multivariate normal distribution with the mean vector and the covariance matrix. Certain regularity assumption needs to be met in mixture model (e.g., Peters & Walker, 1978; Redner & Walker, 1984), a broad modeling family where GMM is rooted in. Interested readers please refer to Muthén and Shedden (1999), Muthén et al., (2002), and Muthén and Asparouhov (2008), for more details of MLE in GMM via EM algorithm.

---

[1] Alternative models could have class-specific parameters in any combinations of factor loadings, covariance matrices of growth factors, and covariance matrices of residuals on the outcome.

Figure 2 illustrates the path diagram of a GMM with 4 times of measurement. The factor loadings $\mathbf{\Lambda}$ are set to be $(\mathbf{1} \quad \mathbf{t})$, where $\mathbf{1}$ is a column of ones and $\mathbf{t} = (0, \quad 1, \quad 2, \quad 3)$, which define the shape of the growth patterns as linear. Assuming the latent classes only differ in their growth mean parameters, latent class (c) does not have pathways to model parameters other than the growth mean parameters. If two latent classes are modeled using the diagram, for example, two distinct class-level growth trajectories will be estimated each with a mean intercept and a mean linear slope, respectively (i.e., $\hat{\alpha}_0^{(1)}, \hat{\alpha}_1^{(1)}$ for latent class 1; $\hat{\alpha}_0^{(2)}, \hat{\alpha}_1^{(2)}$ for latent class 2). Variability estimates of individual intercepts and linear slopes around the mean intercept and mean linear slope parameters are $\hat{\omega}_{\eta_0}$ and $\hat{\omega}_{\eta_1}$, respectively, for each latent class, with their covariance as $\hat{\omega}_{\eta_0\eta_1}$. The residual variance of the outcome is $\hat{\theta}$ at each time point and no correlation is parametrized among the residuals.

This type of GMM is the first-order GMM as the observed outcomes are driven directly by the underlying latent growth factors. In the literature, it is often referred to as GMM where the outcome is a continuous scale score based on multiple items repeatedly measured over time. A sum or a mean score is commonly used to represent the individual's standing on the outcome where the growth trajectories are modeled on.

By using the composite score based on equally weighted items, 1st-order GMM automatically implies that the item-to-scale relationship is the same not only within time but also over time (i.e., satisfaction of L-MI assumption). When the presumption is untenable, as will be shown later, it will lead to fallacious interpretations on the growth characteristics. L-MI assumption calls for an explicit test in order to reach valid conclusions in patterns of change, which is made possible in a SOGMM.

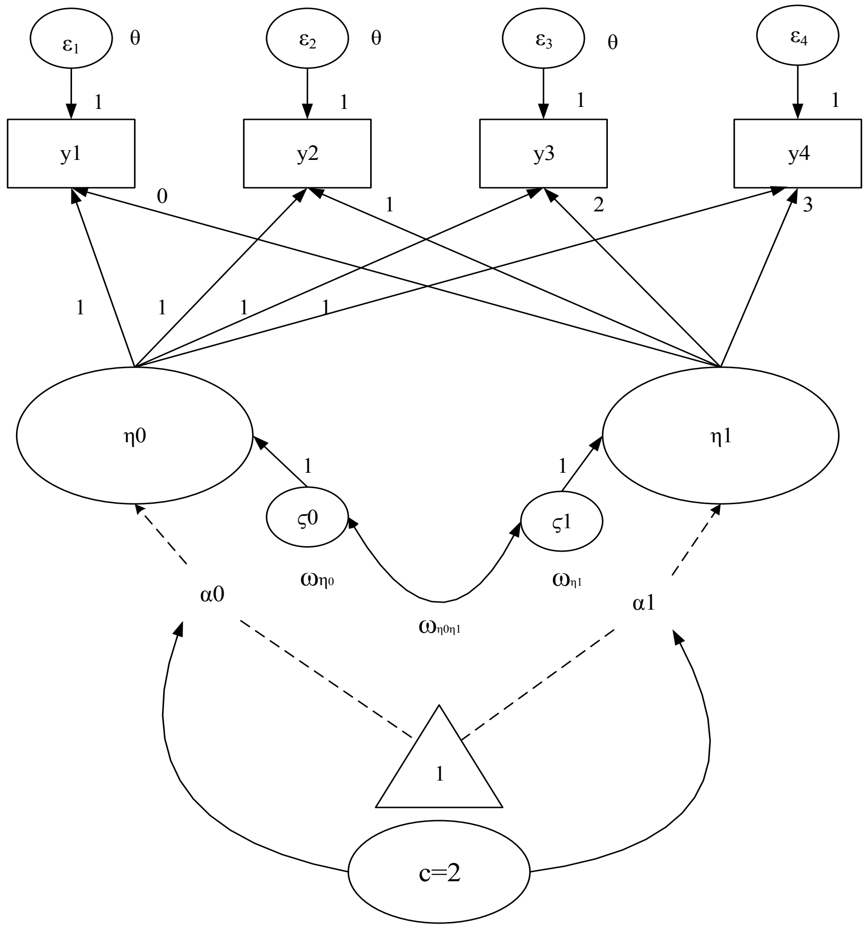

Figure 2. PATH DIAGRAM FOR LINEAR GMM WITH P = 4, Q = 2, C = 2

2.1.2 SOGMM, parameters, interpretation and its advantages over first-order GMM

While first-order GMM implicitly assumes that L-MI is met and the assumption is impossible to be tested without item-level score, SOGMM, on the other hand, allows explicit test for L-MI assumption and simultaneous modeling of longitudinal growth trajectories for multiple latent class. A SOGMM consists of a longitudinal common factor model, a latent growth model, and a mixture component.

The longitudinal common factor model originates from the common factor model (Thurstone, 1947) which defines a common factor as an underlying and unobservable variable that explains the co-variability among observed outcomes. In a common factor model, the observed outcomes are assumed to be a linear function of the underlying common factors and the unique factors (i.e., the part of an observed variable not explained by the common factors) (MacCallum, 2009). As a consequence of this definition, a common factor model partitions the variance of each observed outcome into common variance (i.e., variance accounted for by the common factors) and unique variance (i.e., the variance not accounted for by the common factors).

From SOGMM's perspective, score on each item/indicator is an observed outcome. At each time, the variability in the observed item scores is generated by the same one common factor (or the construct such as depression). The common factor underlying the item scores is called first-order factor. Its relationship to the item scores is captured by the longitudinal common factor model, which serves as the first layer in a SOGMM. Using the relation between first-order factor and observed item score, a longitudinal common factor model can be expressed as

$$x_{jti} = \tau_{jt} + \lambda_{jt}\eta_{ti}^{(c)} + \varepsilon_{jti}, \tag{12}$$

where x_{jti} is the observed score on item j at time t for individual i. τ_{jt} is the item intercept for item j at time t. η_{ti} is the first-order factor scores for individual i at time t. λ_{jt} is first-order factor loadings of item j to the first-order factor at time t, which can be interpreted as latent variable regression coefficients, and ε_{jti} is the unique factor scores for individual i on item j at time t.

A latent growth model serves as the second layer of the SOGMM in order to capture the longitudinal patterns of change. The latent growth factors are used as the 2nd-order factors in explaining the change in the 1st-order common factor. Their relation can be specified as

$$\eta_{ti}^{(c)} = \gamma_0 \xi_{0i}^{(c)} + \gamma_{1t} \xi_{1i}^{(c)} + \varsigma_{ti}, \tag{13}$$

where γ_0 and γ_{1t} are 2nd-order factor loadings, respectively, for relating the 2nd-order factor or the growth factors ($\xi_0^{(c)}$ and $\xi_1^{(c)}$) to the 1st-order factors ($\eta_t^{(c)}$). The 2nd-order factor loadings determine the shape of the growth trajectories. Depending on how the 2nd-order factor loadings are parametrized, the interpretation of the 2nd-order factor can change. When γ_0 is fixed as *1* and γ_{1t} is set as *0*, *1*, …, *t-1*, the 2nd-order factors have the same interpretation as the intercept and linear slope growth factors as in 1st-order GMM. ς_t is the specific factor (or disturbance of the 1st-order common factor) that remains not explained by the underlying growth trajectories (i.e., the 2nd-order factors).

The last layer of SOGMM accounts for the mixture components where 2nd-order growth factors are decomposed of class-specific means and deviations to the means.

$$\xi_{0i}^{(c)} = \alpha_0^{(c)} + \omega_{0i}, \tag{14}$$

$$\xi_{1i}^{(c)} = \alpha_1^{(c)} + \omega_{1i}, \tag{15}$$

where $\alpha_0^{(c)}$ and $\alpha_1^{(c)}$ are the mean intercept and mean linear slope for each latent class, and ω_{0i} and ω_{1i} are the individual random deviations from these means. In SOGMM, as the scale of the first-order factors is arbitrary, the mean of the intercept factor for one of the latent classes needs to be constrained to be 0 for the other means of the intercept factor for the other classes' to be identified.

Using matrix notations, equations (12-15) can be generally written as below. For time *t*,

$$\mathbf{x_t} = \boldsymbol{\tau}_\mathbf{t} + \boldsymbol{\Lambda}_\mathbf{t} \boldsymbol{\eta}_\mathbf{t}^{(\mathbf{c})} + \boldsymbol{\varepsilon}_\mathbf{t}, \ \boldsymbol{\varepsilon}_\mathbf{t} \sim N(\mathbf{0}, \boldsymbol{\Theta}_\mathbf{t}), \tag{16}$$

$$\boldsymbol{\eta}_\mathbf{t}^{(\mathbf{c})} = \boldsymbol{\Gamma}_\mathbf{t} \boldsymbol{\xi}^{(\mathbf{c})} + \boldsymbol{\varsigma}_\mathbf{t}, \ \boldsymbol{\varsigma}_\mathbf{t} \sim N(\mathbf{0}, \boldsymbol{\Psi}_\mathbf{t}), \tag{17}$$

$$\boldsymbol{\xi}^{(\mathbf{c})} = \boldsymbol{\alpha}^{(\mathbf{c})} + \boldsymbol{\omega},\ \boldsymbol{\omega} \sim N(\mathbf{0}, \boldsymbol{\Omega}), \tag{18}$$

where $\mathbf{x_t}$ is a $j \times 1$ vector of observed item score and j is the number of items at time t, $\boldsymbol{\tau}_t$ is a $j \times 1$ vector of item intercepts at time *t*, $\boldsymbol{\Lambda}_t$ is a $j \times r$ matrix of first-order or item factor loadings at time *t*, $\boldsymbol{\eta}_{\mathbf{t}}^{(\mathbf{c})}$ is a vector of $r \times 1$ first-order factor scores for individual *i* who is from latent class *c* at time *t*, and $\boldsymbol{\varepsilon}_{\mathbf{t}}$ is the $j \times 1$ vector of unique factor scores for individual *i* at time t, $\boldsymbol{\Theta}_{\mathbf{t}}$ is the $j \times j$ covariance matrix for the unique factors at time *t*. $\boldsymbol{\Gamma}_t$ is a $r \times s$ second-order factor loading matrix and *s* is the number of 2nd-order factors. $\boldsymbol{\xi}^{(\mathbf{c})}$ is a $s \times 1$ vector of second-order factor scores for individual *i* who is from latent class *c*, and $\boldsymbol{\varsigma}_{\mathbf{t}}$ is a $r \times 1$ vector of latent variable disturbance scores. $\boldsymbol{\alpha}$ is a $s \times 1$ vector of latent growth factor means for latent class *c* and $\boldsymbol{\omega}$ is a $s \times 1$ vector of individual deviations from the means. $\boldsymbol{\Omega}$ is $s \times s$ covariance matrix for the second-order factors, $\boldsymbol{\Psi}_{\mathbf{t}}$ is $r \times r$ covariance matrix for the first-order factors. In SOGMM, as mentioned, since the number of common factor for first order model is 1, $r = 1$ from equation (16) to (17) so that at each time *t*, dimension for the 1st-order factor loading ($\boldsymbol{\Lambda}_{\mathbf{t}}$) and the 2nd-order factor loading ($\boldsymbol{\Gamma}_{\mathbf{t}}$) are respectively, $j \times 1$, and $1 \times s$. The first-order factor score ($\eta_t^{(c)}$), disturbance factor score (ς_t) and covariance for the fist-order factor (Ψ_t) is a scalar (1×1).

Item scores can be expressed in a super-vector as each element in (16) is a sub-vector or sub-matrix in a super-vector or super-matrix (MacCallum, 2009). The observed item scores, item intercepts, and unique factor scores across times can be defined as

$$\mathbf{x} = \begin{vmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_T} \end{vmatrix}_{(J \times T) \times 1} \qquad \boldsymbol{\tau} = \begin{vmatrix} \boldsymbol{\tau}_1 \\ \boldsymbol{\tau}_2 \\ \vdots \\ \boldsymbol{\tau}_{\mathbf{T}} \end{vmatrix}_{(J \times T) \times 1} \qquad \boldsymbol{\varepsilon} = \begin{vmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_{\mathbf{T}} \end{vmatrix}_{(J \times T) \times 1}, \tag{19}$$

where each of the $\mathbf{x_t}$, $\boldsymbol{\tau}_\mathbf{t}$ and $\boldsymbol{\varepsilon}_\mathbf{t}$ are defined as above.

The first-order factor loadings across time, similarly, can be constructed so that

$$\boldsymbol{\Lambda} = \begin{vmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \cdots & \mathbf{0} \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Lambda}_\mathbf{T} \end{vmatrix}_{(J\times T)\times(J\times T)} . \tag{20}$$

The first-order common factors are defined as

$$\boldsymbol{\eta}^{(\mathbf{c})} = \begin{vmatrix} \boldsymbol{\eta}_1^{(\mathbf{c})} \\ \boldsymbol{\eta}_2^{(\mathbf{c})} \\ \vdots \\ \boldsymbol{\eta}_\mathbf{T}^{(\mathbf{c})} \end{vmatrix}_{(r\times T)\times 1} . \tag{21}$$

The unique factor has expected values of 0 and the covariance structure over time is

$$\boldsymbol{\Theta} = \begin{vmatrix} \boldsymbol{\Theta}_{\mathbf{11}} & \boldsymbol{\Theta}_{\mathbf{12}} & \cdots & \boldsymbol{\Theta}_{\mathbf{1T}} \\ \boldsymbol{\Theta}_{\mathbf{21}} & \boldsymbol{\Theta}_{\mathbf{22}} & \cdots & \boldsymbol{\Theta}_{\mathbf{2T}} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Theta}_{\mathbf{T1}} & \boldsymbol{\Theta}_{\mathbf{T2}} & \cdots & \boldsymbol{\Theta}_{\mathbf{TT}} \end{vmatrix}_{(J\times T)\times(J\times T)} , \text{ with } Cov(\boldsymbol{\varepsilon}_\mathbf{t}, \boldsymbol{\varepsilon}_\mathbf{t+n}) = \boldsymbol{\Theta}_\mathbf{t,t+n} . \tag{22}$$

The expected values for disturbance in (17) and random deviations in (18) are zeros so that expected values for the 1st-order common factor scores at each time and across time are respectively,

$$E(\boldsymbol{\eta}_\mathbf{t}^{(\mathbf{c})}) = E(\boldsymbol{\Gamma}_t \boldsymbol{\xi}_\mathbf{i}^{(\mathbf{c})}) = E(\boldsymbol{\Gamma}_t \boldsymbol{\alpha}^{(\mathbf{c})}) = \begin{bmatrix} \gamma_0 & \gamma_{1t} \end{bmatrix} \begin{bmatrix} \alpha_0^{(c)} \\ \alpha_1^{(c)} \end{bmatrix} = \kappa_t^{(c)} , \tag{23}$$

$$E(\boldsymbol{\eta}^{(\mathbf{c})}) = \begin{vmatrix} \kappa_1^{(c)} \\ \kappa_2^{(c)} \\ \vdots \\ \kappa_T^{(c)} \end{vmatrix}_{(r\times T)\times 1} . \tag{24}$$

The covariance structure for the 1st-order common factor at each time is

$$Cov(\boldsymbol{\eta}_{\mathbf{t}}^{(\mathbf{c})}) = \boldsymbol{\Gamma}_{\mathbf{t}}\boldsymbol{\Omega}\boldsymbol{\Gamma}_{\mathbf{t}}' + \boldsymbol{\Psi}, \tag{25}$$

where the variability of the 1st-order common factor at time t is the addition of a scalar (1×1 for $\boldsymbol{\Gamma}_{\mathbf{t}}\boldsymbol{\Omega}\boldsymbol{\Gamma}_{\mathbf{t}}'$) and the time-specific disturbance factor variance with the super-matrix for the covariance for the disturbance factor variance over time as

$$\boldsymbol{\Psi} = \begin{vmatrix} \boldsymbol{\Psi}_{\mathbf{11}} & \boldsymbol{\Psi}_{\mathbf{12}} & \cdots & \boldsymbol{\Psi}_{\mathbf{1T}} \\ \boldsymbol{\Psi}_{\mathbf{21}} & \boldsymbol{\Psi}_{\mathbf{22}} & \cdots & \boldsymbol{\Psi}_{\mathbf{2T}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{\mathbf{T1}} & \boldsymbol{\Psi}_{\mathbf{T2}} & \cdots & \boldsymbol{\Psi}_{\mathbf{TT}} \end{vmatrix}_{(r\times T)\times(r\times T)}, \text{ with } Cov(\boldsymbol{\eta}_{\mathbf{t}}^{(\mathbf{c})}, \boldsymbol{\eta}_{\mathbf{t+n}}^{(\mathbf{c})}) = \boldsymbol{\Psi}_{\mathbf{t,t+n}}. \tag{26}$$

Figure 3 illustrates the path diagram for a SOGMM with responses from 6 items over 4 time points, similar to the 1st-order GMM in Figure 2. In SOGMM, the item scores are used to find the underlying individuals' true level on the construct where the growth trajectories are modeled on, as opposed to 1st-order GMM where observed scale score is used as the outcome directly in modeling growth trajectories. In SOGMM, the bottom longitudinal common factor model captures the first-order common factor (i.e., construct) underlying the item responses at each time while the middle-to-top latent growth model with the mixture component accounts the variability of the construct by using the growth factors as second-order factors. Though both modeling frameworks capture within- and between-individual variability in patterns of change by using the growth mean parameters and the variance estimates of the growth factors, there are benefits in how the model is parametrized in SOGMM.

By using item scores, 1) true representation of individual's standing on the underlying construct can be utilized (Ram & Grimm, 2009), 2) unique variance from the variability on the item scores that is not accounted for by the underlying construct is partitioned out (Harman, 1976). By taking advantage of the 2nd-order factors, variance due to specific factors is separated from the measurement error, resulting in a theoretically error-free measure of the construct

(Chen, Sousa, & West, 2005). More importantly, by using item scores, the relationship between each item and the construct can be assessed explicitly in terms of tenability of longitudinal ME/I, which will be discussed in more details.

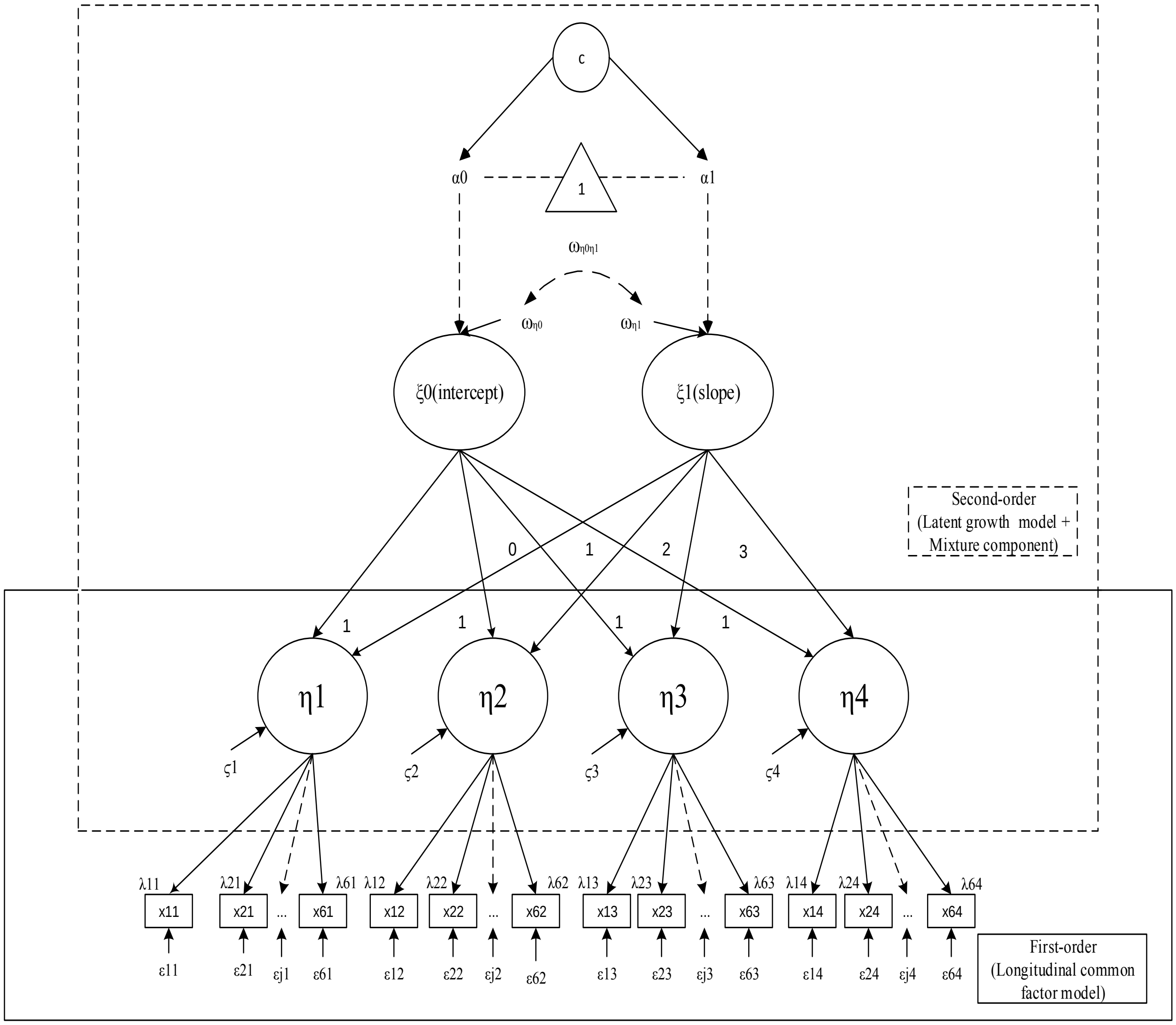

Figure 3. Path diagram of 4 times occasions for a linear SOGMM with *j*=6, *r*=1, *s*=2 (item intercepts omitted)

2.2 LONGITUDINAL MEASUREMENT INVARIANCE (L-MI)

When L-MI is established, meaningful interpretation can be made on whether there is longitudinal change observed on the construct through its manifest variables (i.e., items) (Chan, 1998; Ferrer, Balluerka, & Widaman, 2008; Widaman & Reise, 1997). Different methods can be applied to investigate the L-MI assumption. From the perspective of modeling longitudinal growth patterns, for example, both factorial analytical method (e.g., longitudinal common factor model) and item response theory (IRT) method can be used. The former is of interest in the study as the factorial analytical approach has been utilized in most of the prior studies on longitudinal invariance. Readers interested in using IRT for longitudinal ME/I can refer to papers such as Golembiewski, Billingsley, and Yeager (1976), and Meade and Lautenschlager (2004).

From factorial analytical perspective, MI is considered to be established when a set of model parameters are invariant over different populations where MI is tested. Different levels of MI are defined as whether invariance holds for different sets of item parameters such as item intercepts, factor loadings[2], and unique factor variances (Meredith, 1993). A restricted or nested model with constrains placed on the set of particular item parameters is compared to a nesting model where these parameters are freely estimated. The particular level of MI is said to exist if there is no significant difference in the model fit between the two models (i.e., no significant reduction of model fit using the more parsimonious model). The following are the different levels of MI that are often tested in the literature.

[2] Factor loading in ME/I means factor loadings in the measurement model (i.e., first-order model).

2.2.1 MI levels and their meanings in SOGMM

Complete covariance invariance (Jöreskog, 1971) is considered as an omnibus test assessing the equality of covariance structure over time. When complete covariance invariance is established, the same model-implied covariance structure holds over time. The model under complete covariance invariance assumption is compared to the covariance structure where the item parameters[3] are free to vary over time. If the null hypothesis is retained, it indicates the exchangeability of time (Byrne, Shavelson, & Muthén, 1989). However, if the null hypothesis is rejected, further investigation is required to determine which particular model parameters are statistically different among time.

Configural invariance (Horn, McArdle, & Mason, 1983) (or configurational invariance; Thurstone, 1947; or pattern invariance; Millsap, 1997) is the least restrictive MI assumption following the omnibus test. It aims to assess whether the same factorial pattern holds over time. It requires that the same number of factors with the same patterns of free and fixed factor loadings exist across time. The within-item factor loading among times are not required to be equal to meet this level of invariance. In general, if configural invariance is met, it indicates each time has the same number of factors and that each factor is defined by the same variables (Millsap & Olivera-Aguilar, 2012). Putting in SOGMM framework, it implies that at each time of measurement, the same construct is hypothesized and is defined by the same items. If this level of invariance is rejected, the factorial structure needs to be investigated respectively at each time of measurement in order to justify the instrument and the nature of the construct. On the other hand, once this level of invariance is established, further levels of L-MI should be looked

[3] The parameters are estimated freely for item intercepts, factor loadings, and unique factor variances.

into to locate the source of non-equivalence. However, during estimation, as long as the same factor pattern is extracted, this level of L-MI is maintained.

Metric invariance (Thurstone, 1947) (or weak invariance; Meredith, 1993; or factor pattern invariance; Millsap, 1995) is the next in-line level of L-MI to be tested after establishment of configural invariance (e.g., Widaman & Reise, 1997). It is slightly stringent than configural invariance as metric invariance requires not only the number and pattern of the factor loadings to be the same over time, the factor loadings should also remain the same over time. In SOGMM, it means that there is equality of first-order factor loading for each item among times (i.e., $\mathbf{\Lambda}_t = \mathbf{\Lambda}$). Using the example in Figure 3, metric invariance is met if and only if $\lambda_{j1} = \lambda_{j2} = \lambda_{j3} = \lambda_{j4}$ with $j \forall \in \{1,...,6\}$. Under metric invariance assumption, the 1st-order factor loading of the same item are fixed to be the same over the times while the other parameters (i.e., item intercepts and unique factor variances) are freely estimated.

As the first-order factor loadings are the regression coefficients relating the observed item score to the 1st-order factor (i.e., construct), they can be interpreted as the expected change on the item score with unit change on the 1st-order factor score. Hence, testing whether there is equality on the 1st-order factor loadings among times is essentially assessing equality of scaling units over time. Rejecting the null hypothesis means at least one item has different 1st-order factor loadings at one or more time points. It indicates that the meaning of construct might be different over time and further analysis is needed to find out which items have variant loadings (Olivera-Aguilar, 2013; Widaman & Reise, 1997). Metric invariance has been generally considered the minimum level of MI for the construct to have comparable interpretation (Widaman & Reise, 1997) and it has been examined the most frequently among all MI levels in the published articles in studying

MI (Vandenberg & Lance, 2000). However, some researchers suggested higher level of MI is needed to ensure the same construct is measured over time (e.g., Meredith, 1993).

Strong invariance (Meredith, 1993) is one of the higher levels of MI that could be tested in case when metric invariance is established. It intends to assess whether the observed scores among different time haves the same origin. Strong invariance assumes equality of measurement intercepts in addition to the same factor loadings over time. In SOGMM, this level of L-MI means that there is equality on within-item intercept (i.e., $\boldsymbol{\tau}_{\mathbf{t}} = \boldsymbol{\tau}$) in addition to the equality of within-item 1st-order factor loadings. Using the notations described previously, strong invariance is met if and only if $\lambda_{j1} = \lambda_{j2} = \lambda_{j3} = \lambda_{j4}$ and $\tau_{j1} = \tau_{j2} = \tau_{j3} = \tau_{j4}$ with $j \forall \in \{1,...,6\}$. In addition to the covariance structure traditionally used in testing MI, this level of assumption investigates the model-implied mean structure as well. The argument is that if intercept non-invariance is found, different times have different means on the items and thus individual scores cannot easily be compared across time and items need to be assessed to determine which intercept(s) is/are non-invariant among time. On the other hand, retaining this level of ME/I ensures researchers that the observed difference on the mean level of the items at each time is not due to the measurement.

Strict invariance (Meredith, 1993) (or complete invariance; Millsap, 1995), as another higher level of ME/I, could be assessed after strong invariance is met (e.g., Vandenberg & Lance, 2000). Additional to the within-item equality on item intercepts and first-order factor loadings, within-item unique factor variance should be the same over time (i.e., $\boldsymbol{\Theta}_{\mathbf{t}} = \boldsymbol{\Theta}$) to establish this level of L-MI in SOGMM. In order to test this assumption, the model with equality placed on 1st-order factor loadings, item intercepts and unique factors over time is compared to the model with strong invariance. If there is no difference in model fit, then strict invariance is maintained.

Under strict invariance, the model-implied mean and covariance structure on the observed item scores at time *t* are

$$\mathbf{u}_{\mathbf{x_t}} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\kappa_t^{(c)}, \tag{27}$$

$$\boldsymbol{\Sigma}_{\mathbf{x_t}} = \boldsymbol{\Lambda}\left(\boldsymbol{\Gamma}_{\mathbf{t}}\boldsymbol{\Omega}\boldsymbol{\Gamma}_{\mathbf{t}}^{'} + \boldsymbol{\Psi}\right)\boldsymbol{\Lambda}^{'} + \boldsymbol{\Theta}, \tag{28}$$

where every element is defined previously. If unique factor variance varies with time (i.e., violation of strict invariance), (28) is rewritten as

$$\boldsymbol{\Sigma}_{\mathbf{x_t}} = \boldsymbol{\Lambda}\left(\boldsymbol{\Gamma}_{\mathbf{t}}\boldsymbol{\Omega}\boldsymbol{\Gamma}_{\mathbf{t}}^{'} + \boldsymbol{\Psi}\right)\boldsymbol{\Lambda}^{'} + \boldsymbol{\Theta}_{\mathbf{t}}, \tag{29}$$

where the variability on the observed item scores is different over time and the difference could be due to non-invariant unique factor variance or changing second-order factor loadings over time. To make sure that the observed changes in the mean and covariance structure over time can be only attributed to changes in the latent variable, not the variability, of the measurement error not accounted by the latent construct, researchers need to test strict invariance. This level of L-MI is important as it increases both the confidence and validity of research findings (DeShon, 2004; Lubke & Dolan, 2003; Meredith, 1993).

As shown, there are a number of levels of L-MI that could be tested in SOGMM. With increasing strictness of the assumption (i.e., higher level of L-MI), more sets of item parameters are considered invariant over time so that more confidence could be put in interpreting the observed changes without the influence of measurement. As indicated, in order to test the level L-MI, sequential tests need to be performed. In literature, there is no universal consensus as the specific order of the levels of MI to test (Bollen, 1989; Steenkamp & Baumgartner, 1998). However, it is generally recognized there is a hierarchical order among the tests such as starting from configural and followed by metric and strong invariance (Vandenberg & Lance, 2000).

2.2.2 Partial L-MI

The MI levels described are said to maintain if and only if when the whole set of model parameters remains the same over time, which might be too stringent to obtain. This leads to the less stringent MI conditions what is called partial MI. Partial MI (Byrne et al., 1989; Reise, Widaman, & Pugh, 1993) in a longitudinal sense is defined as only a subset of the model parameters maintaining their equality over time instead of the whole set.

Partial L-MI can be tested when researchers reject the hypothesis of a specific level of L-MI as described earlier. Using the example from Figure 3, if metric invariance is rejected, for instance, partial metric invariance could be tested for fixing $\lambda_{j1} = \lambda_{j2} = \lambda_{j3} = \lambda_{j4} \;\; j \;\; \forall \in \{1,...,4\}$ while allowing the other two item's 1st-order factor loadings free to vary over time. The partial metric model can then be compared to configural invariance in their model fit. If the null hypothesis is retained, partial metric invariance is established. Most of the existing literature examined partial MI when metric invariance is rejected but it can happen after rejecting any level of MI.

Compared to the sequential tests for the level of MI, there has been more debate on partial MI (e.g., Byrne et al., 1989; Millsap & Hartog, 1988).With the rejection of a specific level of MI, in order to test partial MI, a choice must be made on the selection of reference indicators (e.g., items) whose item parameter values are supposedly to be invariant over time. The items whose parameter values are non-invariant over time are considered as offending items and hence need to be freely estimated. A sequence of tests will start from comparing the nested model holding constant item parameters on the reference indicators and freeing offending parameters to the nesting model where a level of MI was rejected. The choice of reference indicator(s) could influence the item parameters' estimates on subsequent tests of invariance, and thus the conclusion about MI pattern could change drastically (Ferrer et al., 2008). It calls for guide of theory and contextual knowledge in addition to the empirical statistical results when choosing the reference indicators. However, it is out of the scope of this paper and more in-depth discussion on the topic can be found somewhere else (e.g., Byrne et al., 1989; Steinmetz, 2011; van de Schoot et al., 2013). In general, compared to the MI assumptions (i.e., the levels described in 2.1.3), partial MI is more practical and makes it possible to assess the changes on the outcome which otherwise might not be appropriate (Vandenberg & Lance, 2000).

2.2.3 Identification and identification invariance

There are infinite numbers of estimates of the item parameters that can be found to produce the same mean and covariance structure. In order to find unique solution, identification constraints are needed. One of the two alternatives is often adopted to identify covariance structure, each with their consequences. Put in SOGMM, one way is to fix one particular item's 1^{st}-order factor loading to be 1 at the initial time and subsequent times. It results in 1) the 1^{st}-order factor to be on the same scale with the same item constrained and 2) the variance of the 1^{st}-order factor to be freely estimated. The other alternative is to fix the 1^{st}-order factor variance at 1 so that the factor being measured will be on a standardized normal scale (i.e., mean of 0 and standard deviation of 1) and the 1^{st}-order factor loadings will be estimated freely. This is not appropriate in growth model as change is of interest. When the item used for identification is truly invariant, it is said to maintain identification invariance. Depending on whether the item constrained meets identification invariance, it could impact the interpretation in modeling growth (Wirth, 2008).

2.3 MEASUREMENT IN LATENT GROWTH

The current practice in estimating growth in GMM context is using composite such as mean score of a scale. Reliability such as Cronbach's alpha is calculated at each time to support the use of these scale scores. However, high reliability index does not justify or guarantee the satisfaction of L-MI assumption. With violation of L-MI, growth parameters and model fit can be biased (Leite, 2007; Olivera-Aguilar, 2013; Wirth, 2008). The following section illustrates what happens in modeling growth when there is such violation. Examples have been demonstrated by Leite (2007), Wirth (2008), and Olivera-Aguilar (2013) in LGM which can be considered as a special case of GMM when there is one subpopulation (i.e., c=1).

Wirth (2008) used both mean score and factor score as indicators in capturing growth characteristics where the performance of factor scores using a number of methods was compared to the mean score. He demonstrated factor score's ability in recovering growth parameters for a linear trend when high level of L-MI is maintained. Over 4 times of measurement, the relationship between a mean score (y_{ti}) and the underlying growth trajectory can be specified as

$$\begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \eta_0 \\ \eta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{bmatrix}, \qquad (30)$$

where the mean score alternatively could be a factor score with item parameters in a longitudinal common factor model as $y_{ti} = J^{-1}\left(\sum_{j=1}^{J}(\tau_{jt} + \lambda_{jt}\eta_{ti} + \varepsilon_{jti})\right)$, leading to rewriting the above equation as

$$\begin{bmatrix} J^{-1}\left(\sum_{j=1}^{J}(\tau_{j1}+\lambda_{j1}\eta_{1i}+\varepsilon_{j1i})\right) \\ J^{-1}\left(\sum_{j=1}^{J}(\tau_{j2}+\lambda_{j2}\eta_{2i}+\varepsilon_{j2i})\right) \\ J^{-1}\left(\sum_{j=1}^{J}(\tau_{j3}+\lambda_{j3}\eta_{3i}+\varepsilon_{j3i})\right) \\ J^{-1}\left(\sum_{j=1}^{J}(\tau_{j4}+\lambda_{j4}\eta_{4i}+\varepsilon_{j4i})\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \xi_{0i} \\ \xi_{1i} \end{bmatrix} + \begin{bmatrix} \varsigma_{1i} \\ \varsigma_{2i} \\ \varsigma_{3i} \\ \varsigma_{4i} \end{bmatrix}, \tag{31}$$

assuming it is adequate to use mean score represent the true score on the construct. If we further assume that all measurement intercepts equal to 0, factor loadings equal to 1, and average of the residual score over items is 0 for all time, the above equation could be re-written as a series of linear equations as

$$\begin{bmatrix} 4^{-1}(0+4\eta_{1i}+0) \\ 4^{-1}(0+4\eta_{2i}+0) \\ 4^{-1}(0+4\eta_{3i}+0) \\ 4^{-1}(0+4\eta_{4i}+0) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \xi_{0i} \\ \xi_{1i} \end{bmatrix} + \begin{bmatrix} \varsigma_{1i} \\ \varsigma_{2i} \\ \varsigma_{3i} \\ \varsigma_{4i} \end{bmatrix}, \tag{32}$$

so that

$$\begin{aligned} \eta_{1i} &= (\xi_{0i}+\varsigma_{1i}) \\ \eta_{2i} &= (\xi_{0i}+\xi_{1i}+\varsigma_{2i}) \\ \eta_{3i} &= (\xi_{0i}+2\xi_{1i}+\varsigma_{3i}) \\ \eta_{4i} &= (\xi_{0i}+3\xi_{1i}+\varsigma_{4i}) \end{aligned}. \tag{33}$$

The equations in (33) indicate that when strict invariance holds, the invariant item intercepts and factor loadings do not have a systematic effect on the growth parameters such that $E(\eta_t)=\Gamma_t\xi+\varsigma$.

On the other hand, when there is configural invariance, using the values from Leite (2007)'s example, as the item intercepts and factor loadings are,

$$\tau' = [0 \quad 0 \quad 0 \quad 0 \quad .25 \quad .25 \quad .25 \quad .25 \quad .5 \quad .5 \quad .5 \quad .5 \quad .75 \quad .75 \quad .75 \quad .75], \tag{34}$$

$$
\Lambda' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.5 & 1.5 & 1.5 & 1.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.5 & 2.5 & 2.5 & 2.5 \end{bmatrix}, \tag{35}
$$

the linear equation is now

$$
\begin{bmatrix} 4^{-1}(0+4\eta_{1i}+0) \\ 4^{-1}(1+6\eta_{2i}+0) \\ 4^{-1}(2+8\eta_{3i}+0) \\ 4^{-1}(3+10\eta_{4i}+0) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \xi_{0i} \\ \xi_{1i} \end{bmatrix} + \begin{bmatrix} \varsigma_{1i} \\ \varsigma_{2i} \\ \varsigma_{3i} \\ \varsigma_{4i} \end{bmatrix}. \tag{36}
$$

The latent factor $\boldsymbol{\eta}_\mathbf{t}$ can be solved as

$$
\begin{aligned}
\eta_{1i} &= (\xi_{0i} + \varsigma_{1i}) \\
\eta_{2i} &= 0.67(\xi_{0i} + \xi_{1i} + \varsigma_{2i} - .25) \\
\eta_{3i} &= 0.50(\xi_{0i} + 2\xi_{1i} + \varsigma_{3i} - .50) \\
\eta_{4i} &= 0.40(\xi_{0i} + 3\xi_{1i} + \varsigma_{4i} - .75)
\end{aligned}. \tag{37}
$$

Equation (37) shows that in the presence of systematically changing measurement characteristics over time, the relationship between the latent construct and the underlying growth factors are differentially weighted. This not only indicates that the measurement characteristics (e.g., factor loading) have different effects at each measurement occasion so that $E(\eta_t) \neq \Gamma_t \xi + \varsigma$, but also implies that using mean score replacing the true score on the construct in modeling the growth trajectories is inappropriate when there is violation of L-MI.

2.4 SUMMARY OF STUDIES IN L-MI IN LGM

Based on the above example, it can be seen that using mean score as indicator in modeling growth is likely to produce biased parameter estimates when there is violation of L-MI. Effect of violation of longitudinal measurement non-invariance has been studied by a few researchers in LGM where one average growth trend is sufficient to describe the inter-individual differences in individual growth trajectories.

Leite (2007) studied and compared effects of violation of L-MI in curve-of-factor model (i.e., 2nd-order LGM[4]; McArdle, 1988) and LGM using mean score (i.e., 1st order LGM) using Monte Carlo simulation. In his study, the independent variables (IVs) systematically manipulated included number of time (3, 5), number of items (5, 10, 15), total sample size (100, 200, 500, 1000), type of item (congeneric, essentially tau-equivalent), level of ME/I (strict, metric, and configural), and mean reliability for the mean composite (.7, .9). Model fit indices were examined along with growth parameters including means of intercept and slope factors, their variances, covariance, and standard errors.

In general, growth parameters were estimated more accurately when the item-to-construct relation was accommodated by the 2nd-order LGM. LGM using mean composite produced consistently larger bias on the growth parameter estimates under same violation of L-MI compared to the 2nd-order LGM. . With a higher degree of violation, the degree of relative bias was more severe. When each item score had an additive and multiplicative constant in its relation to the true score (i.e., congeneric items), the absolute magnitude of relative bias, for example, can be .49 and .47 for variance of intercept, variance of slope, respectively. Both numbers were reduced to .02 when the violation of L-MI lessened to the case where all items were essentially tau-equivalent. When equivalence of within-item factor loading over time can be assumed (i.e., essentially tau-equivalent), almost all growth parameters were able to be recovered with

---

[4] As contrast to the factor-of-curve model which is also a 2nd-order LGM which does not address the growth of a single latent construct measured by multiple indicators, but the common growth of multiple observed variables.

negligible relative bias using mean score as outcome in LGM. The only exception is the mean of the intercept factor mean which was estimated with unacceptable large bias regardless of whether the items were congeneric and even essentially-tau equivalent. Compared to LGM using mean composite, the performance of 2nd-order LGM in recovering the accuracy in growth parameters was robust to violation of L-MI. Its ability to produce unbiased growth parameter estimates was not influenced by the type of item and L-MI deviation. The relative bias magnitude for all growth parameters estimates was negligible using 2nd-order LGM.

Though 2nd-order LGM was found better at recovering parameter estimates than LGM using mean score, the study was associated with a few limitations that restrict the ability in its generalization. There was no systematic change in the measurement characteristics over time manipulated in the study as the within-item factor loadings and item intercepts were selected randomly between .5 to 1, and 0 and 1, respectively. Item parameters included in the study assumed either strict, metric or configural invariance excluding the possibility of examination on partial L-MI. Additionally, there was no specification as which and whether the same item was constrained for identification and whether there was identification invariance on the same item over time. Without such information, the question arises as what scale is the latent construct on at each time and whether, longitudinally, they are comparable. It made the examination of the results of the study somewhat difficult.

Wirth (2008) conducted a simulation study in assessing longitudinal measurement non-invariance in LGM context and included conditions to address some of the questions that Leite (2007)'s study did not answer. Systematic changes on item parameters were simulated in order to study the impact of the measurement characteristics on the growth model parameters. Longitudinal measurement non-invariance was generated on either the item intercepts or on both of item intercepts and factor loadings. More specifically, a decreasing or a mixed changing pattern was simulated for these parameters when there was measurement non-invariance. A decreasing pattern was defined as decreasing parameter values on the same item over time whereas a mixed pattern was defined at between-item level where some items parameters' values increased and the others decreased over time. Partial L-MI was also incorporated in examining the recovery of the growth parameters as the items were either 1) strictly maintaining specific measurement invariance (8 out of 8 items), 2) partially invariant (2 out of 8 items) or 3) non-invariant (0 out of 8 items). Identification of the scale on the construct used either one or all items (i.e., assuming L-MI on one item or on all items). As a consequence, for some conditions where the item truly invariant was constrained for identification, identification invariance was maintained whereas, for others, it was not. Other factors were sample size (250, 750), item-specific, time adjacent unique factor correlations (0, .1), factor score methods (mean score, regression-based factor scores, 2^{nd}-order LGM-based factor scores), and trajectory shape (linear, free-loading). Evaluation criteria included bias, relative bias, and RMSE of growth parameters, model fit indices and the likelihood of rejecting the true functional form.

By incorporating the systematic changes on item intercepts and factor loadings, relationship between degree and direction of bias on growth parameter estimates and generating item intercepts and factor loadings were discovered. In general, regardless of the scoring methods and the scores used in modeling growth, for fixed effects (i.e., intercept factor mean and slope factor mean), the degree of bias was found to be correspondent to the difference of 1) the (mean) constrained generating factor loading and unity, and 2) the (mean) constrained generating item intercept and 0. For random effects including variance of intercept factor, variance of slope factor, and their covariance, the degree of bias was only influenced by the difference in the (mean) constrained generating factor loading and 1.

As for direction of the bias, when using mean score, it was the direction that the differences themselves might reflect. For example, the mean intercept estimate was negatively biased when the generating mean intercept and mean factor loading were less than 0 and 1. As the author indicated, when using mean score, the presence of bias was the result of the discrepancy between the true measurement model and the equal weighting mechanism adopted by the model using mean score. As the gap between the two got bigger, parameter estimates with larger bias were expected. As a consequence, it was expected that the consistent decreasing pattern of measurement characteristics included in the study produced larger biases than those from Leite (2007)'s study as the generating item intercepts and factor loadings were systematically deviated from 0 and 1.

Contrast to the similar impact from the decreasing pattern, mixed pattern non-invariance exhibited differential influence on the growth parameters. Intercept factor mean estimates resulted from using mean score were associated with severe bias and highest RMSE when there was a mixed pattern on either item intercepts or on both item intercepts and factor loadings. Slope factor mean estimates, on the other hand, had negligible bias when non-invariance was only on item intercepts using mean score as indicator for latent growth but large bias and RMSE when both item intercepts and factor loadings had mixed changing patterns. Under non-invariance mixed pattern on item intercepts, both intercept and slope factor variance had raw bias around 0 but their RMSE (including the RMSE on their covariance) were substantially inflated. When both item intercepts and factor loadings had mixed non-invariant patterns, the random effects recovery was worse.

Correlated measurement error was associated with producing less well recovered intercept, slope factor variance and their covariance with inflated relative bias and RMSE but this factor had negligible influence on the means for intercept and slope factors. Sample size was found to not influence recovery of the growth parameters, in comparison to the other factors in the study, but rather the model fit indices and the rejection rate of the functional form of the average growth trajectory.

Partial L-MI with identification invariance and factor score methods were found to produce unbiased growth factor parameters. When there were no items maintaining measurement invariance including the single item constrained for identification, both the fixed and random effects were not accurately recovered regardless of the scoring methods (mean or factor scores). However, when there was partial measurement invariance, factor score methods, in particularly, using factor score by constraining single item produced least amount of bias in both fixed effects and random effects, which was better than recovering growth parameters than using the mean score.

Olivera-Aguilar (2013) took a step further in studying longitudinal measurement non-invariance by including other systematic measurement characteristics in LGM including total number of items (6, 9, 15), the proportion of non-invariant items (1/3, 2/3), and degree of L-MI violation. Bias, relative bias, standard errors of the parameter estimates, their RMSE, and model fit were examined. The L-MI deviation was simulated either on factor loadings fixing item intercepts and unique factor scores (i.e., non-invariant item intercepts) or on item intercepts fixing factor loadings and unique factor scores (i.e., non-invariant item factor loadings). Contrary to Wirth (2008) who included decreasing within-item/mixed between-item pattern on item parameters when there was measurement non-invariance, factor loadings and item intercepts were simulated to have decreasing pattern only and increasing pattern only, respectively. Both patterns were created within items. Sample size was set at 100, 200, 500, or 1000.

When sum scores were used as LGM outcomes, degree of growth parameters bias and relative bias were found expectedly increasing with higher degree of contamination (% of items exhibiting non-invariance). Both non-invariant intercepts, and non-invariant factor loadings were associated with biased growth parameters. More specifically, both of the non-invariant conditions affected slope factor mean recovery and non-invariant factor loadings conditions additionally negatively influenced recovery of slope factor variance and covariance between the intercept and slope factor.

Though it was of hope to provide more insight of unaccounted measurement non-invariance in LGM, Olivera-Aguilar (2013)'s simulation design is limited in using only the sum score as outcome variables in LGM whereas the data was generated using curve-of-factors model. All growth parameters were found not recovered well even in conditions with invariant item intercepts and factor loadings. The generating parameter values were adjusted after the simulation study to avoid inflated bias, relative bias and RMSE. It was done, separately, for conditions with different number of items so that with measurement invariance, under each different total number of items condition, there was no bias for all growth parameters. This method however, comes with some costs. With various population parameters for different number of items, comparisons cannot be made between the effects from different numbers of items on parameter recovery. Similar logic could be applied to the other factors that were originally designed to be crossed with the number of items. The fact that the other parameters, such as unique factor variance, had unadjusted (i.e., incorrect) values restricts the generalizability on the results of the study. More importantly, the question still remains open in regards to the effects of the factors manipulated such as the contamination level on the growth parameter recovery as the parameters might be adjusted overly.

2.5 CONCLUSION

There has been a rise in studying growth in longitudinal setting where various growth-associated questions get answered such as rate of growth and individual variability on the growth rate. However, there has been less attention paid to the role of measurement invariance in the growth context while L-MI is a prerequisite to make sure there is no systematic change due to the measurement over time in drawing valid conclusion on the longitudinal changes in the underlying construct.

Effects of violation of L-MI have been studied in a handful of studies in LGM framework on growth parameter recovery. In general, with higher degree of violation of L-MI (including increased magnitude, more non-invariant items, and fewer sets item parameters exhibiting measurement invariance over time), LGM using composite scores produced growth parameters with large bias. In contrast, second-order LGM model or using factor score as indicator with satisfactory of identification invariance in modeling growth was found to be more robust in recovering the growth parameters. Though LGM is a very useful tool in examining the growth pattern in a longitudinal setting, it assumes that there is homogeneous population underlying the growth pattern. This assumption might be untenable in some situations where multiple subpopulations exist.

GMM accounts for heterogeneity in the population, addresses questions such as average rate of growth for each latent class and classification of individuals to a latent class associated with certain growth characteristics. However, longitudinal measurement non-invariance has not been well understood in a mixture of subgroups that associate with different growth patterns. Though SOGMM is available in studying longitudinal patterns of change for latent classes and can be used to test explicitly L-MI, it has not been studied with the presence of longitudinal measurement non-invariance. As a matter of fact, up to date, to our knowledge, there is no study systematically examining the violation of L-MI in GMM context.

3.0 METHODOLOGY

The purpose of the study is to evaluate how the longitudinal measurement non-invariance impacts the growth factor recovery and subject classification in GMM using first- and second-order GMMs. The general research questions include:

1) Are growth factors in GMM affected when longitudinal measurement non-invariance is present? If yes, what factors contribute to the biased/imprecise growth parameters?
2) Does the L-MI deviation impact the classification accuracy and if yes, what factors contribute to poor classification accuracy?
3) Do the factors affect the growth parameters in the same way with the same magnitude for each latent class?

A Monte Carlo study was used to answer these questions. This chapter is organized into eight sections; 1) Literature review of GMM in empirical studies, 2) fixed simulation factors, 3) manipulated simulation factors, 4) generating model, 5) data generation, 6) data analysis, 7) evaluation criteria, and 8) data generation validation.

3.1 LITERATURE REVIEW OF GMM APPLICATION

A literature review on GMM was performed to help determine the configurations in the simulation. Fifty studies were extracted from PsycINFO database (10/25/2014) by using the keyword of “Growth Mixture Model”. L-MI was assessed in only 2 (4%) of the 50 studies (Callina, Johnson, Buckingham, & Lerner, 2014; Kirves, Kinnunen, De Cuyper, & Mäkikangas, 2014) where the rest used first-order GMM with a composite score. About half of the time, the items measuring the construct was on a continuous scale. The score on the instrument was either based on the responses collected from all the items on an instrument (e.g., Lutz et al., 2014; Schumm, Walter, & Chard, 2013) or a subset of the items from a more general instrument or assessment. For example, Zerwas, Von Holle, Watson, Gottfredson, and Bulik (2014) selected items related to childhood anxiety from Child Behavior Check List (CBCL/ age 4-18 years) that measure anxious feelings and anxiety-related somatic symptoms; Callina et al., (2014) used six items from the Search Institute’s Profiles of Student Life-Attitudes and Behaviors (PSL-AB) questionnaire to measure trust, defined by adolescences’ perceived positive connections with their parents. Among the studies, the number of repeated measures ranged from 2 to 12. Most often, the studies were based on 4 occasion of measurement (14; 28%). Linear growth pattern was discovered for majorities of the studies (32; 64%). Total number of latent classes discovered varied from 2 to 7. Twenty-eight studies reported the specific estimator they used. Of these 28 articles, 26 used ML estimator. The growth factor estimates could vary largely depending on the field of research and the instrument used. For the studies which specified the growth parameter estimates and their standard errors, 95% confidence interval (CI) was constructed between paired growth factor mean estimates from any two latent classes. The results indicated the configurations of (dis)similarity in growth factor means among the latent classes could vary largely. For example, the intercept factor and the linear slope factor could have statistically similar means among classes (e.g., Galatzer-Levy, 2011; Lydecker, 2011; Thompson, Swartout, & Koss, 2013) (i.e., CI contained a difference of 0). In some other cases, both the intercept and linear slope factor could have statistically different means among the classes (e.g., Hong et al., 2014; Liu, 2012; Mustillo, Hendrix, & Schafer, 2012). It can also be a combination of the above two situations where intercept and slope factor means among some latent classes were statistically different while they were not among the others (e.g., Gunn et al., 2013; Lavender et al., 2013). Lastly, the total sample size as well as the size for the smallest class could vary largely (See Table 1).

Table 1. Descriptive statistics for total sample size and sample size for the smallest class

	Mean	SD	Min	Max
Total Sample Size	1175	1568	88	10099
Sample Size for the smallest class	155	385	3	2191

3.2 FIXED FACTORS

The section summarizes the factors held constant across conditions. Given there has been no research so far investigated L-MI violation in GMM, the number of latent classes was set as 2 to allow easy interpretation. Given the literature review in GMM above, time of measurement was fixed at 4 with linear trend for both of the latent classes. Mixture proportions were set at .8 and .2 representing the proportion of individuals from one dominant class and one rare class. Total sample size was fixed at 1000 so that the sample sizes for the latent classes were representative of the average sample sizes found in empirical research. The total number of items was fixed at 6. The number was chosen for two reasons. Firstly, the number of items was within the range of the item numbers examined by Leite (2007) and Olivera-Aguilar (2013). Additionally, as indicated from literature, a smaller number of items often occurred in practice as the items were chosen from a larger instrument catering to the operational definition of a specific construct. No missing data was generated in the simulation to keep the conditions under the study manageable.

3.3 MANIPULATED FACTORS

The current study varied six factors including 1) number of items violating L-MI, 2) combination of item intercepts (i.e., τ) and item loadings (i.e., λ), 3) absence or presence of within-item correlated measurement error, 4) direction of change on non-invariant item intercepts, 5) latent class distance, and 6) model used to estimate growth factor parameters and individual classification. This section describes and justifies the specific levels of each factor included in the study.

3.3.1 Contamination level of measurement non-invariance

The factor was expressed as percentages of a set of item parameters violating L-MI. Three contamination levels were considered: 0%, 50%, and 100%. 0% of the items violating L-MI was defined as all item parameters on the 6 items exhibiting a certain level of L-MI. When the contamination was more than 0%, there were either 3 items (50%) or 6 items (100%) having a set of non-invariant item parameters. Under 50% contamination, non-invariance was simulated on item intercepts only, item loadings only, or on both item intercepts and item loadings. Under 100% contamination, either all item loadings were non-invariant or all item intercepts were non-invariant except the reference item's intercept.

Both Wirth (2008) and Olivera-Aguilar (2013) varied the number of items exhibiting measurement non-invariance in examining the impact of L-MI violation on the growth parameter recovery. Zero, two or all eight items had non-invariant item parameters in Wirth (2008)'s study. However, contamination level was not included as a factor in the analysis in examining its systematic effect on the growth factor recovery. The proportion of non-invariant items was set either as 1/3 or 2/3 in Olivera-Aguilar (2013), resulting in 2 to 10 non-invariant items out of a total of 6, 9, or 15 items simulated. This study fixed the total number of items at 6 with an intention to mimic more severe contamination level associated with smaller number of items. The different contamination levels included would provide answer to the research question as whether higher level contamination is associated with more biased/less precisely estimated growth factor and classification rates.

3.3.2 Longitudinal measurement non-invariance pattern

Different generating L-MI patterns were created by combination of item loadings (λ) and item intercepts (τ) in order to assess the impact of varying degree of measurement non-invariance on growth parameters and classification rates recovery. The combinations under the study generally included 1) strict, 2) strong, and 3) partial invariance. Particularly, under strict invariance, item loadings and item intercepts were combined to form the pattern of invariant loadings and invariant intercepts on all items in addition to invariant unique factor variance. Under strong invariance, item loadings and intercepts were invariant but the unique factor variance was different on the same item over time. Under partial invariance, with 50% contamination, L-MI patterns were created as invariant loadings on all items and partially invariant intercepts on 3 items ( ), partially invariant loadings on 3 items and invariant intercepts on all items

($pLiT$), and partially invariant loadings and partially invariant intercepts on 3 items ($pLpT$). With 100% contamination, more severer L-MI deviation patterns were included as partially invariant loadings on 3 items and non-invariant intercepts on all but the identification item ($pLNiT$), and non-invariant loadings on all items and partially invariant intercepts on 3 items ($NiLpT$).

Different combinations of the item loadings and item intercepts represented various L-MI deviation configurations which were studied by Leite (2007), Wirth (2008) and Olivera-Aguilar (2013). Item parameters were simulated to meet strict, metric or configural invariance in Leite (2007). Olivera-Aguilar (2013) included 1) non-invariant item loadings while keeping item intercepts and unique factor variances the same and 2) non-invariant item intercepts while maintaining metric invariance on item loadings. Wirth (2008), on the other hand, simulated a wide range of combinations including 1) maintenance of strong/metric invariance, 2) full or partial non-invariance on item intercepts under metric invariance, 3) partial invariance on both item intercepts and loadings, and 4) non-invariance on both sets of item parameters for all items.

In general, the combinations included in the study were to replicate the conditions included in Wirth (2008)'s study but with modifications. In the current study, the combinations of item loadings and item intercepts are associated with different levels of contamination. Specifically, $iLiT$ pattern was under 0% contamination which included both the strict and strong L-MI patterns. Patterns of $iLpT$, $pLiT$ and $pLpT$ designated the partial invariance patterns associated with 50% contamination. Patterns of $pLNiT$ and $NiLpT$, on the other hand, resembled escalated L-MI violation where most of the item parameters were non-invariant. Identification invariance on the referent item was maintained on both item loading and intercept under $pLNiT$ while $NiLpT$ violated identification invariance where none of the item loadings were invariant.

The specific combinations of item intercept and item loading were created by considering 1) whether each set of the item parameters, respectively, was invariant or not, and 2) if non-invariant, whether it was fully non-invariant, or partially non-invariant. Non-invariant condition on both item intercept and loading ($NiLNiT$) was not included in the current study as it was found by Wirth (2008) not able to produce accurate growth factor estimates regardless of the analysis model and the item(s) used for identification.

In practice, when an item is found invariant over time, it is constrained to have an item loading of one across all occasions for identification purpose. This item is named as the identification item or referent item. The action of constraining an identification item's loading which is truly invariant to be the same over time results in identification invariance where the model is essentially correctly specified (Hancock, 2005).

The current study assumed the presence of identification invariance items. With identification invariance, at least one item needed to be identified as invariant and the same item(s) was/were constrained for identification purpose over time. As a consequence, it resulted in 7 combinations of item loadings and item intercepts out of 9 possible combined patterns under different contamination levels (combination of iL , pL , or NiL paired with iT , pT , or NiT). The spectrum of the combinations would be able to provide information on the modeled outcomes when 1) both item loadings and item intercepts were invariant, 2) only item loadings were invariant, 3) only item intercepts were invariant, and 4) varying degree of measurement non-invariance with minimum partial L-MI maintained.

All the combinations created under the second-order growth factor had different unique factor variance except under strict invariance. No partial invariance condition was generated under strict invariance. The purpose of including this specific level of invariance was to establish a “golden-standard” in assessing recovery of growth factors and classification rate. Under strict L-MI assumption, observed change on the outcomes can be only attributed to the underlying construct (i.e., growth factor scores) instead of to any shift on measurement characteristics over time. Due to the fact that performance of SOGMM is unknown, though this level of invariance might be hard to meet in research practice, it was considered necessary to be included in the simulation.

3.3.3 Within-item correlated measurement error over time

Presence of within-item correlated measurement error was defined as a constant time-adjacent unique factor correlation within the same item. Within-item correlated measurement error on unique factors was simulated to be either present or absent. When present, the correlated measurement error was fixed at .1 for the same items over time. This value replicated Wirth (2008)’s study. The magnitude was small but it was found to systematically influence the growth factor recovery when first-order factor scores and second-order growth factor scores were estimated separately in a 2-step procedure. This study used the same magnitude of correlated measurement error and intended to examine its effect when the factor scores are estimated simultaneously altogether.

3.3.4 Directional change on the non-invariant item intercepts

Item intercepts were either consistently increasing or decreasing when they were non-invariant. In contrast, non-invariant item loadings were simulated with a decreasing pattern in the current study. The decreasing pattern on the item loadings was consistent with the changing pattern on the item loadings in Wirth (2008) and Olivera-Aguilar (2013). Both of these studies suggested that the bias direction on the slope factor mean estimates is related with the intercept change direction. However, item intercepts with opposite change direction has not been examined as Olivera-Aguilar (2013) included only increasing pattern and Wirth (2008) investigated within-item decreasing pattern and between-item mixed pattern for item intercepts. This factor aimed to address the question whether the direction in the bias on the slope factor mean corresponds to the direction on non-invariant item intercepts.

As the non-invariant item intercepts were found to produce biased estimate on the slope factor mean (Olivera-Aguilar, 2013), the two patterns simulated were expected to provide answer to their effect on the bias direction in the slope factor means for the latent classes. The simulated direction was used as a superscript with conjunction of specific generating L-MI pattern in reporting the results. For example, for $iLNiT$ level, $iLNiT^{+}$ indicates all non-invariant item intercepts had an increasing pattern while $iLNiT^{-}$ means a decreasing pattern for the non-invariant item intercepts, with all item loadings invariant over time.

3.3.5 Latent class distance

Two levels of latent class distance were included in the study. Mahalanobis distance (MD) was used to indicate how distinguishable the latent classes were. MD was set as either 5 or 1.5 for extremely or intermediately well separated latent classes. Extremely large effect sizes of latent class distance were found in empirical research (e.g., Allan et al., 2014; Yaroslavsky, Pettit, Lewinsohn, Seeley, & Roberts, 2013). In simulation studies (Depaoli, 2013; Liu, 2012), indistinguishable latent classes were found to adversely influence growth factor recovery in first-order GMM. Thus, different latent class distances are needed to examine the impact of L-MI violation on growth characteristics recovery respectively under distinguishable and harder-to-distinguish latent classes.

Parameters such as within-class variability could be varied to reflect different levels of MD, but most often the growth factor means have been manipulated to study the effect of latent class distance in the literature (e.g., Depaoli, 2013; Liu, 2012; Peugh & Fan, 2012). Hence, growth factor means, in particular the slope factor means for the rare class, were varied to produce different levels of MD. The values of the growth factors were specified in the following section describing the generating model.

3.3.6 Analysis model

Different analysis models were included to examine their comparative performance in GMM context with L-MI deviation. The generated data was analyzed, respectively, by 3 GMMs including the 1st-order GMM using mean score, a constrained SOGMM and a freely estimated SOGMM. In the constrained SOGMM, all item parameters including those for the referent item were constrained to be equal over time. In the freely estimated SOGMM, all item parameters but those for the referent item were freely estimated over time.

Both Leite (2007) and Wirth (2008) included different analysis models when there was longitudinal measurement non-invariance in LGM. Growth factor parameters produced by 2nd-

order model (or using factor scores) were found with less bias than first-order model (or using observed scores). Both growth factor means and variability estimates were found to recover well as long as identification invariance was maintained (Wirth, 2008).

This study is primarily interested in finding the optimal model in the presence of L-MI violation. Among the analysis models included, the constrained SOGMM had the advantage of appropriately accounting for covariance structure among items in contrast to the first-order model. The freely estimated SOGMM, on the other hand, was better than the constrained SOGMM at additionally accounting for non-invariant item parameters. As a consequence, the growth factor parameters were expected to be recovered well by the freely estimated SOGMM when identification invariance holds. The constrained SOGMM might perform well in the second place while the first-order GMM would be mostly impacted by the L-MI deviation. On the other hand, given the prevalence of first-order models, the study intended to assess conditions under which the 1^{st}-order GMM was still safe to use in interpreting the class-specific growth characteristics and classification rate, if such conditions existed.

3.3.7 Summary of the factors and their conditions

The factors manipulated in the simulation were not fully crossed with each other. Table 2 illustrates the design conditions formed by the factors and their levels, with each combination of item loading and item intercept multiplied by the number specified under each factor, resulting in a total of 132 conditions. For ease of reporting, Table 3 summarized each of the factors with their corresponding levels.

Table 2. Design conditions for the study

Contamination	Combination of Loading and Intercept	Directional Change on Item Intercepts	Correlated Measurement Error	Latent class distance	Analysis model	Total
0%	strict	None	x 2	x 2	x 3	12
	strong	None	x 2	x 2	x 3	12
50.00%	iLpT	x 2	x 2	x2	x 3	24
	pLiT	None	x 2	x 2	x 3	12
	pLpT	x 2	x 2	x 2	x 3	24
100.00%	pLNiT	x 2	x 2	x 2	x 3	24
	NiLpT	x 2	x 2	x 2	x 3	24
					Total	**132**

Table 3. Acronyms for the study conditions to be used in presentation of results

Factor	Acronym	Definition
Contamination	0%	Full L-MI where all items maintained a specific level of L-MI
	50%	Partial L-MI where 3 items have non-invariant item parameters
	100%	All items had non-invariant item parameters
Combination of λ and τ	*iLiT*	Invariant λ and invariant τ on all items
	iLpT	Invariant λ on all items and non-invariant τ on 3 items
	pLiT	Non-invariant λ on 3 items and invariant τ on all items
	pLpT	Non-invariant λ and non-invariant τ on 3 items
	pLNiT	Non-invariant λ on 3 items and non-invariant τ on all but referent item
	NiLpT	Non-invariant λ on all items and non-invariant τ on 3 items
Within-item correlated measurement error	No correlated ME	Correlated measurement error over time for the same item was 0.
	Correlated ME	Correlated measurement error over time for the same item was .1.
Directional change on item intercepts	None	Item intercepts were invariant over time.
	+	Non-invariant item intercepts increased over time.
	–	Non-invariant item intercepts decreased over time.
Latent class Distance	$MD = 5$	Latent classes are separately extremely well.
	$MD = 1.5$	Latent classes are separately with intermediate distance.
Analysis model	C2-GMM	Data was analyzed using constrained SOGMM.
	F2-GMM	Data was analyzed using freely estimated SOGMM.
	1-GMM	Data was analyzed using 1st-order GMM on mean score composite.

3.4 GENERATING MODELS

The generating model was a SOGMM corresponding to 2 latent classes whose linear growth trajectories were defined on 6 items over 4 time points. The growth factor parameter values in this study were set in general to mimic prior simulation studies in which individuals were the same at the baseline between the latent classes and then became more distinct over time. With this setup, this study focused on the influence of L-MI deviation on the rate of change rather than the initial status. Specifically, population growth parameter values were set to mimic the study from Muthén, Asparouhov, and Nylund (2007) who investigated the performance of enumeration indices (e.g., AIC, BIC) in correctly choosing the number of latent classes in different mixture models. The population parameters are summarized in Table 4.

Table 4. Population growth factor values for the two latent classes

Model parameters	Dominant Class (80%)	Rare Class (20%)
Growth parameters		
Intercept factor mean	0	0
Var (Intercept factor)	0.25	0.25
Slope factor mean	0	1.00 (or .2)
Var (Slope factor)	0.04	0.04
Covar (Intercept, Slope)	0.02	0.02

The variances of the intercept factor and the slope factor were the same for the two classes. The intercept and slope factors were correlated with a correlation of .2 within each class. Intercept factor means were simulated as the same between the two latent classes. The mean on the slope factor for the rare class was set as either 1 or .2 to reflect large or small latent class distances, with all the other growth factors held to be the same under varying MD conditions.

The disturbance variance was chosen so that 80% of the variability in the first-order factor score was explained by the growth factors at each time. According to equation (17), the variance of the 1st-order factor scores[5] and the proportion of the variance accounted for by the growth factors were, respectively,

$$Var(\eta_t) = \gamma_t \Omega \gamma_t' + \psi = \omega_{\eta_0} + \gamma_t^2 \times \omega_{\eta_1} + 2\gamma_t \omega_{\eta_0\eta_1} + \psi \,, \tag{38}$$

$$R^2(\eta_t) = \frac{\omega_{\eta_0} + \gamma_t^2 \times \omega_{\eta_1} + 2\gamma_t \omega_{\eta_0\eta_1}}{\omega_{\eta_0} + \gamma_t^2 \times \omega_{\eta_1} + 2\gamma_t \omega_{\eta_0\eta_1} + \psi} \,. \tag{39}$$

Based on the parameter values in Table 4 and the above formula, the resulting calculated disturbance variance values were listed in Table 5.

Table 5. Generating disturbance variance across conditions

Time	Variance η_t	Variance ς_t
1	0.31	0.06
2	0.41	0.08
3	0.61	0.12
4	0.91	0.18

3.4.1 Item parameters

Item parameters (loading and intercept) were based on the measure on the lack of self-efficacy in poly-substance use for the group of norm adolescence (Pentz & Chou,1994). Self-efficacy in treating adolescent misuse of substance is an important construct as it was found to predict substance use outcome even after controlling for treatment or intervention (Ramo, Anderson, Tate, & Brown, 2005). Adolescence is a period of time where fast development takes place in

[5] Latent class membership was omitted since the 2nd-order factor loadings and growth factor covariance structure was assumed to be the same between latent classes.

cognitive, physical, behavioral, social and emotional domains. It is important to make sure the youth perceive, interpret and endorse the items measuring self-efficacy the same way over the course of time in order to conclude with more informative etiology model in studying the developmental path of self-efficacy, finding out potential risk and protective factors, and devising proper interventions.

The following section describes the generating values for item loadings and item intercepts. The L-MI patterns were created by combing the specific values of the item loading and the item intercept. Item parameters were set to be the same for the dominant and rare classes to maintain the between-class measurement invariance. Under such setting, dominant class on average had a flat growth trajectory resembling stable self-efficacy throughout adolescence. In contrast, the rare class was depicted by an increased average growth indicating elevated risks with substance use during this time period. The general patterns (whether stable or increasing) as mentioned previously were reflected in the slope factor means, respectively, for the two latent classes.

Table 6 had the generating values for item factor loadings. For the iL (invariant condition), the loadings were generated so that the average of the factor loadings equaled to 1 at each time point. For the pL condition, 3 items (50% items) exhibited changing loadings on the same items. For the NiL condition, factor loadings on all items including the referent item were non-invariant over time. For pL and NiL conditions, the item factor loading at the first time was the same as that for iL condition. The last three items were defined to have non-invariant factor loadings under partial invariant condition. When there were non-invariant factor loadings, only the pattern of decreasing was simulated, and the decrease had a uniform amount of .08 between adjacent time points. Across 4 occasions, the overall decrease on the factor loading for each item was .24, which was similar to the magnitude in Wirth (2008)'s study. Olivera-Aguilar (2013) defined .2 as a small effect size on the change of item factor loading.

Table 7 had the specific generating values for the item intercepts when there was increasing non-invariant trend and Table 8 summarized the values for the item intercepts when there was a decreasing non-invariant trend over time. The generating average values for the intercept for the iT condition were controlled at 1.9. The same item intercepts at time 1 were used across different L-MI patterns (iT, pT, NiT) at initial time of measurement. The last three items were set to have non-invariant item intercepts under partial invariant condition. With non-invariance, there was a consistent amount of .14 applied to the change in the same item intercept between the adjacent time points. The overall change over time (.42) were also similar to the magnitude of the change on item intercept in Wirth (2008)'s study. The change of the direction was the opposite between Table 7 and Table 8 on items with non-invariant item intercepts but the magnitude of change was the same.

All items were constrained when fitting the constrained SOGMM and the first item was constrained for identification when fitting the freely estimated SOGMM. The item parameters for the 1st item were generated as truly invariant except for the non-invariant factor loading condition (NiL). Due to the fact that the condition of $NiLNiT$ was not included in the study, not all item parameters were non-invariant simultaneously. That is to say, at least partial measurement invariance was maintained on either the item factor loadings or item intercepts.

Table 6. Population parameters of factor loadings for various L-MI patterns

Condition	Contamination	Item	Time 1	Time 2	Time 3	Time 4
iL	0.00%	**1**	**1**	**1**	**1**	**1**
		2	1	1	1	1
		3	1.027	1.027	1.027	1.027
		4	0.846	0.846	0.846	0.846
		5	0.846	0.846	0.846	0.846
		6	1.3	1.3	1.3	1.3
		average	1	1	1	1
pL	50.00%	**1**	**1**	**1**	**1**	**1**
		2	1	1	1	1
		3	1.027	1.053	1.053	1.053
		4	*0.846*	*0.766*	*0.746*	*0.726*
		5	*0.846*	*0.766*	*0.746*	*0.726*
		6	1.3	*1.22*	*1.2*	*1.18*
		average	1	0.97	0.96	0.95
NiL	100.00%	***1***	***1***	***0.92***	***0.9***	***0.88***
		2	*1*	*0.92*	*0.9*	*0.88*
		3	*1.027*	*0.947*	*0.927*	*0.907*
		4	*0.846*	*0.766*	*0.746*	*0.726*
		5	*0.846*	*0.766*	*0.746*	*0.726*
		6	*1.3*	*1.22*	*1.2*	*1.18*
		average	1	0.923	0.903	0.883

Note: Item 1 is the referent item (Bold) in freely estimated SOGMM; non-invariant items parameters are italicized

Table 7. Population parameters of item intercepts with increasing L-MI pattern

Condition	Contamination	Item	Time 1	Time 2	Time 3	Time 4
iT	0.00%	**1**	**1.813**	**1.813**	**1.813**	**1.813**
		2	1.813	1.813	1.813	1.813
		3	1.884	1.884	1.884	1.884
		4	1.684	1.684	1.684	1.684
		5	1.684	1.684	1.684	1.684
		6	2.5	2.5	2.5	2.5
		average	1.9	1.9	1.9	1.9
pT+	50.00%	1	**1.813**	**1.813**	**1.813**	**1.813**
		2	1.813	1.813	1.813	1.813
		3	1.884	1.884	1.884	1.884
		4	*1.684*	*1.824*	*1.964*	*2.104*
		5	*1.684*	*1.824*	*1.964*	*2.104*
		6	*2.5*	*2.64*	*2.78*	*2.92*
		average	1.9	1.97	2.04	2.11
NiT+	100.00%	**1**	**1.813**	**1.813**	**1.813**	**1.813**
		2	*1.813*	*1.953*	*2.093*	*2.233*
		3	*1.884*	*2.024*	*2.164*	*2.304*
		4	1.684	*1.824*	*1.964*	*2.104*
		5	1.684	*1.824*	*1.964*	*2.104*
		6	*2.5*	*2.64*	*2.78*	*2.92*
		average	1.9	2.01	2.13	2.25

Note: Item 1 is the reference indicator (Bold) in freely estimated SOGMM; non-invariant items parameters are italicized.

Table 8. Population parameters of item intercepts with decreasing L-MI pattern

			Time			
Condition	Contamination	Item	1	2	3	4
iT	0.00%	**1**	**1.813**	**1.813**	**1.813**	**1.813**
		2	1.813	1.813	1.813	1.813
		3	1.884	1.884	1.884	1.884
		4	1.684	1.684	1.684	1.684
		5	1.684	1.684	1.684	1.684
		6	2.5	2.5	2.5	2.5
		average	1.9	1.9	1.9	1.9
pT-	50.00%	1	**1.813**	**1.813**	**1.813**	**1.813**
		2	1.813	1.813	1.813	1.813
		3	1.884	1.884	1.884	1.884
		4	*1.684*	*1.544*	*1.404*	*1.264*
		5	1.684	*1.544*	*1.404*	*1.264*
		6	*2.5*	*2.36*	*2.22*	*2.08*
		average	1.9	1.83	1.76	1.69
NiT-	100.00%	**1**	**1.813**	**1.813**	**1.813**	**1.813**
		2	*1.813*	*1.673*	*1.533*	*1.393*
		3	*1.884*	*1.744*	*1.604*	*1.464*
		4	*1.684*	*1.544*	*1.404*	*1.264*
		5	*1.684*	*1.544*	*1.404*	*1.264*
		6	*2.5*	*2.36*	*2.22*	*2.08*
		average	1.9	*1.78*	*1.66*	*1.55*

Note: Item 1 is the reference indicator (Bold) in freely estimated SOGMM; non-invariant items parameters are italicized.

3.4.2 Unique factor variance

Time-specific unique factor variance is defined differently over varying combinations of τ and λ. Under strict invariance, unique factor variance equals to .2 for all items at each time point, resulting with item communality ranging from .53 to .88. Under all other invariance conditions with different levels of contamination on factor loadings, unique factor variance at different times for each item is listed in Table 9 where an increment of .05 was simulated for each item. These unique factor variance values resulted in item communalities ranging from .40 to .81 under different levels of non-invariance on factor loadings.

Table 9. Generating unique factor variance for conditions other than strict invariance

Item	time 1	time 2	time 3	time 4
1	0.3	0.35	0.4	0.45
2	0.2	0.25	0.3	0.35
3	0.4	0.45	0.5	0.55
4	0.2	0.25	0.3	0.35
5	0.3	0.35	0.4	0.45
6	0.2	0.25	0.3	0.35

3.5 DATA GENERATION

The data was generated in SAS 9.4 for each latent class, separately. Individual 2nd-order factor scores, disturbance factor scores and unique factor scores were generated from univariate normal distributions, respectively, with dimension of $n \times 2$, $n \times 4$, and $n \times 24$ where n was the sample size for the specific latent class. They were then converted to multi-normal distributed random variables so that their mean and covariance structure equaled to the parameters described in previous sections. For the 2nd-order factor scores, the expected value, respectively, for the dominant and rare class was

$$E(\xi^1) = (0 \quad 0) \qquad E(\xi^2) = (0 \quad 1), \tag{40}$$

with the common covariance structure as

$$Cov(\xi^1) = Cov(\xi^2) = \begin{pmatrix} .25 & .02 \\ .02 & ,04 \end{pmatrix}. \tag{41}$$

For the disturbance factor scores, its multivariate normal distributions after transformation was

$$\varsigma \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} .06 & 0 & 0 & 0 \\ 0 & .08 & 0 & 0 \\ 0 & 0 & .12 & 0 \\ 0 & 0 & 0 & .18 \end{bmatrix} \right). \tag{42}$$

The unique factor scores were generated from a multivariate normal distribution with a 24×1 mean vector of 0 and covariance of

$$\Theta = \begin{vmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} & \Theta_{14} \\ \Theta_{21} & \Theta_{22} & \Theta_{23} & \Theta_{24} \\ \Theta_{31} & \Theta_{32} & \Theta_{33} & \Theta_{34} \\ \Theta_{41} & \Theta_{42} & \Theta_{43} & \Theta_{44} \end{vmatrix}_{(6\times4)\times(6\times4)}, \tag{43}$$

with diagonal elements for the covariance of unique factors among the items at each time, and off-diagonal elements for the covariance structure of the correlated unique factors among the items over time. The diagonal elements were the same for each time and it can be specified as follows

$$\Theta_{11} = \Theta_{22} = \Theta_{33} = \Theta_{44} = \begin{pmatrix} .2 & 0 & 0 & 0 & 0 & 0 \\ 0 & .2 & 0 & 0 & 0 & 0 \\ 0 & 0 & .2 & 0 & 0 & 0 \\ 0 & 0 & 0 & .2 & 0 & 0 \\ 0 & 0 & 0 & 0 & .2 & 0 \\ 0 & 0 & 0 & 0 & 0 & .2 \end{pmatrix}. \quad (44)$$

When there was no correlated measurement error ($\theta^{0}_{jT,j(T+n)}$), all the other values in the super matrix in (43) was 0. With the presence of the correlated measurement error ($\theta^{1}_{jT,j(T+n)}$), all the other off-diagonal matrices in (43) were fixed as

$$\Theta_{jt,j(T+n)} = \begin{pmatrix} .02 & 0 & 0 & 0 & 0 & 0 \\ 0 & .02 & 0 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 & 0 \\ 0 & 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & 0 & .02 & 0 \\ 0 & 0 & 0 & 0 & 0 & .02 \end{pmatrix}. \quad (45)$$

After transforming the 2nd-order factor scores and disturbance scores, first-order factor scores for each individual on each item were created by using equation (17). Individual item scores were then formed by combining the generating values of λ and τ from Table 6 to Table 8 with the first-order factor scores and generated unique factor scores. Composite scores were calculated as the mean of item scores at each time. Datasets generated separately for individual latent classes were combined together as the final dataset for analysis. Individual item scores and the composite scores were used as input variables for Mplus 7 to fit the SOGMMs and the 1st-order GMM, respectively.

3.6 DATA ANALYSIS

When fitting SOGMMs, each individual item score was used as indicator of the lack in self-efficacy construct on which the growth trajectories were modeled. When the data was analyzed using SOGMMs, regardless of the conditions, for the common factor model at each time point, 1) unique factor scores and the within-item correlation over time were estimate freely, 2) the item intercept for the first item was constrained to be the same over time and 3) the factor loading on the first item was fixed as 1. For the constrained SOGMM, for the rest of the items, item factor loadings and item intercepts were constrained equal over the four time points. In other words, identification invariance was assumed on all item parameters in this analysis model. The freely estimated SOGMM, on the other hand, placed identification invariance assumption only on the referent item. With the rest of the item parameters being freely estimated, conditional on whether the referent item was truly invariant, measurement non-invariance might be appropriately accounted for. Thus, undue influence from violation of L-MI could be minimized on the underlying growth factor estimates. Number of latent classes to extract was specified as 2. The 2nd-order factor loadings were specified as 0, 1, 2, and 3 to impose linear shape for the growth trajectories for both latent classes. For the growth factors, no constraints were put on them so that their mean and (co)variances were estimated freely[6].

When fitting 1st-order model, mean scores averaged over the items at each time period were used as outcome for which the growth trajectories were modeled. Similar to fitting the SOGMMs, 2 latent classes were extracted with their average growth trajectory patterns identified

[6] Except the means on the intercept factors for both latent classes so that influence on particularly the slope factors can be studied.

as linear (i.e., factor loadings of 0, 1, 2, 3). Moreover, growth factors were freely estimated for both latent classes except the means on the intercept factor.

During estimation of the models, a maximum of 1000 iterations were allowed for model convergence. After a number of pilot tests, the numbers of starting values in the initial and final optimization stages were set as 100 and 20 to avoid problems in non-convergence. A replication was saved once it converged properly to a global solution. Number of replications with non-convergence (e.g., not-replicable best likelihood, negative variance estimate, inadmissible solution) was recorded. More than 900 replications were allowed in order to reach minimally 200 replications in each design cell. The first 200 replications which converged properly were retained for analysis in each cell.

In each of the analysis models of any generated dataset, besides convergence rate, we recorded 1) estimate of each latent class's slope factor mean and variance and 2) overall and class-specific classification among individuals.

3.7 EVALUATION CRITERIA

Evaluation criteria of this simulation study included convergence rate, raw bias, relative bias, and root mean square error (RMSE) of the slope factor estimates. Raw bias was used to evaluate recovery of class-specific slope factor means. It was defined as the difference between the estimate and its corresponding population parameter. For each of the slope factor means, raw bias was calculated as

$$B\left(\hat{\theta}_{rd}^{(c)}\right) = \hat{\theta}_{rd}^{(c)} - \theta_{d}^{(c)}, \tag{46}$$

where r is the r^{th} replication, $\hat{\theta}_{rd}^{(c)}$ is estimate of the slope factor mean of latent class *c* for replication *r* in condition *d* and $\theta_{d}^{(c)}$ is the population generating value of latent class *c* in condition *d.* As the slope factor mean for the dominant class was simulated as zero, use of raw bias on both slope factor means enabled fair comparisons between the latent classes.

Relative bias was used to assess recovery of class-specific slope factor variance. It was calculated as the ratio of raw bias over the corresponding population parameter using the notations defined previously.

$$\mathrm{ReB}(\hat{\theta}_{rd}^{(c)}) = \frac{\mathrm{B}\left(\hat{\theta}_{rd}^{(c)}\right)}{\theta_{d}^{(c)}}. \tag{47}$$

RMSE is a measure of the variability of the slope factor mean and variance estimates over the number of replications (i.e., 200). It was defined as the square root of the average squared difference between the parameter estimate and its population value.

$$\mathrm{RMSE}(\hat{\theta}_{rd}^{(c)}) = \left[\mathrm{R}^{-1}\sum_{r=1}^{R}(\hat{\theta}_{rd}^{(c)} - \theta_{d}^{(c)})^{2}\right]^{\frac{1}{2}}. \tag{48}$$

Analysis of variances (ANOVAs) was performed under each MD condition to determine the effect of the simulation factors on overall/class-specific classification accuracy, and raw bias and relative bias for slope factor mean and variance. The separated analysis for conditions with these two different class distances enabled examination of other simulation factors on the outcomes, avoiding the dominant impact from large differences simulated in the latent class distances. All the other factors were treated as between-subject factors since each of the final 200 replications was generated differently under each level of the generating condition and the replications saved in each cell were different among the three analysis models.

Directional change on item intercepts was nested under generating L-MI pattern, which was in turn nested under contamination level. Analysis model and within-item correlated measurement error were crossed with each other and the rest of the factors in the design. Full model with all the main effects with up to 3-way interactions were examined. Effects with partial Eta-squared larger than .02 was considered for reporting.

The findings on the recovery in the growth rates were augmented with empirical Type I error and power rates. Type I error was examined on the slope factor mean for the dominant class (C1) for which the slope factor mean parameter was generated as zero. Percent of replications within each cell rejecting the null hypothesis was defined as Type I error rate. This calculation was restricted to generating condition of strong invariance with presence of correlated ME as this resembled the model specification in the 2nd-order models where the unique factor variances and correlated ME were freely estimated. Such restriction avoided the influence from false measurement models and consequently, the parameters were expected to follow the sampling distribution under the null hypothesis. Power, on the other hand, was assessed on the slope factor mean for the rare class (C2) where the true parameter value was non-zero. Percent of replications within each cell that correctly rejecting the null hypothesis was calculated. In contrast to calculation of Type I error, empirical power rates were examined across varying generating conditions to gauge the impact of deviation of measurement assumption and the robustness of the analysis model under such deviation.

3.8 DATA GENEARTION VALIDATION

A subset data was generated for verification. Due to the fact that the data was generated class by class, validation was performed on class-specific data first. Then the two classes' data was pooled together for further validation. The purposes of validating the data included examination of whether 1) a generated L-MI can be appropriately identified, and 2) item parameters and growth parameters can be recovered.

For purpose one, the data was generated for one replication under strict invariance and $iLpT^{+}(33.33\%)$ with absence of correlated measurement error. Strict invariance was chosen given it is the most optimal measurement situation. Under strict invariance, within each latent class, the observed difference on the outcome scores can be attributed only to latent variable other than the measurement property. If the data was generated as intended, it was expected to have no raw bias on the fixed effects (e.g., growth factor means particularly slope factor means) and negligible raw bias on the random effects (e.g., slope factor variance). As most of the conditions to be created presented measurement non-invariance, $iLpT^{+}(33.33\%)$ was chosen representing the less optimal measurement conditions. It was looked into as whether partial invariance could be detected.

Chi-square difference test, RMSEA, CFI, and SRMR were used to determine whether the generated invariance level was met. A non-significant *p*-value from Chi-square difference test, absolute change on RMSEA, CFI, and SRMR being less than .02, .01, .03 (Chen, 2007), respectively, indicated no difference in model fit. When data was generated under strict invariance, all the model fit indices indicated the strict invariance assumption was met. Moreover, it was able to detect partial strong invariance properly. Using the same sample generated under strict invariance, both item parameters and growth parameter estimates were examined. There were negligible differences between the true parameter values and their corresponding estimates.

4.0 RESULTS

The study aimed to examine the influence of varying L-MI deviation configurations on growth rate recovery and classification accuracy, and performance of alternative growth mixture models. This chapter is organized in four sections. Section 1 reports convergence rates over all simulation conditions. Section 2 reports results for the simulation conditions with large class separation (MD=5) with a focus on the following outcomes: 1) relative/raw bias and RMSE of growth factor means and variances, 2) empirical Type I error and power rates on class-specific slope factor mean, and 3) overall/class-specific classification accuracy. Section 3 reports results on the same outcomes for the simulation conditions with small class separation (MD=1.5). Section 4 provides a summary and comparison on results from the two latent class distances.

In the sections examining the accuracy (relative/raw bias and RMSE) of growth rates, raw bias was used for both slope factor means allowing comparison between latent classes. Relative bias was used for assessing recovery of slope factor variances. Raw and relative biases with a magnitude greater than .05 were considered practically significant. For recovery of growth rates, results started with partially nested ANOVAs to assess the effects of various simulation factors. Partially nested ANOVAs were used because there was a nested structure among the simulation design factors of item intercepts change, generating L-MI pattern, and contamination level. Modelled effects included main effects, 2-way interactions, and 3-way interactions. Effects with at least small sizes ($\eta_p^2 > .02$) were interpreted. Descriptive results were then collapsed among the levels of factors which had negligible effect sizes ($\eta_p^2 \le .02$). Recovery of growth rates were accompanied by the results on the empirical Type I error and power, with Type I error on the slope factor mean in the dominant class under conditions with L-MI assumed, and power on the slope factor mean in the rare class under all generating L-MI configurations. Lastly, recovery of classification was assessed by the count of individuals classified into each latent class (i.e., mixture proportion in unit of sample size) and the count of individuals into their generating latent class using partially nested ANOVAs. The counts enabled examination of classification accuracy at class- and individual-level, respectively.

4.1 CONVERGENCE

A replication was defined as convergent if it reached to a proper global solution with no report of negative variance estimates, local maximum or inadmissible solution. Convergence rate was calculated as the percentage of replications that converged out of the 900 replications per design cell. Convergence rates were associated with latent class separation within the scope of this study. Replications converged 100% times under extremely large latent class separation (MD=5) while the convergence rates varied with the increased overlap between the latent classes (MD=1.5) (Table 10). With intermediate distance between the latent classes, convergence rates ranged from 27% to 44% with similar convergence rates among varying levels of manipulated factors except the analysis models. The marginal convergence rate was increased by 33% using the F2-GMM as compared to the other two analysis models.

Convergence rates in this study were found to be low under MD = 1.5. Low convergence rates have been documented for complicated GMMs such as in Liu and Hancock (2014). Hence, the low convergence rates were expected for the 2nd-order GMMs (i.e., C2-GMM and F2-GMM) where item parameters were estimated, and both of 1st- and 2nd-order factors were scored. For 1-GMM, the low convergence rates might be due to the problems in estimating growth factors resulted from harder-to-separate latent classes. This would be discussed in more detail in the following section 4.3. Given the low convergence rate, 200 converged replications per design cell were selected for further analysis.

Table 10. Convergence rates by analysis models under each generating condition under MD=1.5.

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Intercept Change	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	29%	32%	38%	33%	31%	43%
0%	strong	None	31%	33%	42%	30%	31%	43%
50%	pLiT	None	29%	32%	41%	31%	29%	38%
50%	iLpT	+	31%	34%	41%	31%	32%	42%
50%	pLpT	+	33%	32%	41%	34%	29%	40%
100%	pLNiT	+	33%	33%	39%	29%	34%	41%
100%	NiLpT	+	31%	30%	37%	30%	31%	37%
50%	iLpT	-	29%	30%	39%	31%	30%	40%
50%	pLpT	-	28%	27%	41%	30%	28%	44%
100%	pLNiT	-	32%	32%	43%	33%	30%	40%
100%	NiLpT	-	28%	32%	42%	32%	28%	39%

4.2 GROWTH CHARACTERISTICS RECOVERY UNDER MD = 5

Table 11 presented the factors with significant effect sizes in the ANOVAs. Effects with negligible sizes were not reported, including 3-way interactions, main effect of correlated ME and its 2-way interactions with other factors. Symbols indicating the nesting structure between factors were omitted to save space.

In general, the accuracy in the growth parameters recovery was mostly explained by contamination %, generating MI pattern, analysis model and their interactions. The accuracy in the recovery of the slope factor mean also depended on item intercepts change. The factors with significant prediction of the relative/raw bias were, to some extent, similar between the two latent classes. More specifically, the factors significantly predicting the bias in slope mean are similar in C1 and C2, while the latter has dissipating effect sizes for nearly all factors. The exception was the main effects of contamination % and generating MI pattern which had larger effect sizes in C2 than in C1.

4.2.1 Slope factor mean

4.2.1.1 Raw bias

Based on the ANOVA results, contamination %, generating MI pattern, item intercepts change, analysis model and their 2-way interactions significantly impacted the raw bias of the slope factor means for both latent classes. Mean raw bias under combinations of generating MI pattern and item intercepts change were plotted for each analysis model in *Figure 4*, separately for C1 and C2.

The effect of MI pattern, item intercepts change and analysis model interacted in interpreting raw bias on the slope factor means. When there was no change on the item intercepts, slope factor mean was estimated accurately in C1 regardless of the analysis model. In C2, however, estimates on the slope factor mean were positively biased using 1-GMM, but not C2-GMM or F2-GMM.

Table 11. ANOVA results investigating class-specific growth rate estimates accuracy when MD = 5.

Latent Class	Parameter Estimate	**Main Effects**				**2-way Interactions**		
		Model	Contamination	MI Pattern	Intercepts Change	Model × Contamination %	Model × MI Pattern	Model × Intercepts Change
C1	Slope Mean*	0.85	0.08	0.05	0.83	0.16	0.09	0.71
	Slope Variance	0.44	0.03	0.04	-----	0.09	0.08	-----
C2	Slope Mean*	0.44	0.19	0.17	0.41	0.03	0.05	0.27
	Slope Variance	0.19	-----	-----	-----	-----	-----	-----

Note: * indicates raw bias was used on slope factor mean as outcome in the ANOVA instead of relative bias.

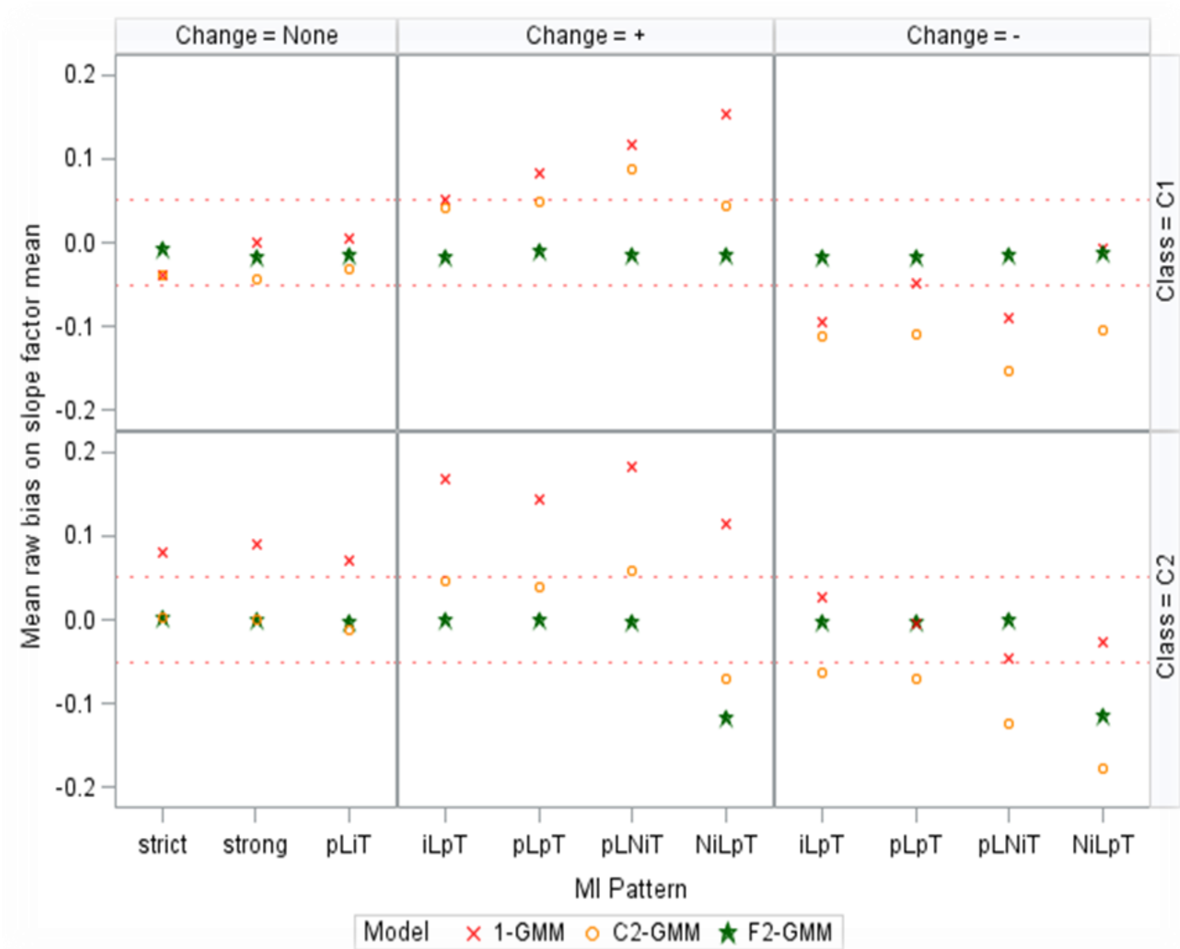

Figure 4. Mean raw bias on class-specific slope factor mean under combinations of MI pattern and item intercepts change for each analysis model when MD = 5. Line of |.05| was drawn as reference.

With increased item intercepts, raw bias on slope factor means were negligible using F2-GMM under all generating L-MI patterns except under NiLpT condition in C2. C2-GMM produced marginally acceptable raw bias on the average growth rate under the same generating L-MI pattern. 1-GMM was only able to produce accurate slope factor mean for the dominant class (C1) under iLpT. In C1, under the rest of the generating L-MI patterns, raw bias consistently increased with larger deviation from L-MI assumption. The raw bias was unacceptable in C2 regardless of generating L-MI pattern, with less distinctive pattern in the change of magnitude of raw bias compared to that in C1.

When item intercepts decreased, F2-GMM performed the best among the analysis models. For F2-GMM, the calculated mean raw bias exhibited the same pattern, compared to when item intercepts increased. Performance of C2-GMM was again not as good as F2-GMM in either latent class, just as in the condition with increased intercepts. The slope factor means were not estimated accurately by C2-GMM with mean raw bias larger than |.05|. Moreover, mean raw bias produced by C2-GMM was larger in magnitude under the same MI generating pattern with decreased item intercepts in contrast to increased item intercepts. 1-GMM, on the other hand, performed better than C2-GMM but not as well as F2-GMM. More specifically, the mean raw bias produced by 1-GMM was (marginally) acceptable among the various generating L-MI patterns.

Besides magnitude on raw bias, direction on raw bias of slope factor mean estimate was influenced by the directional change on item intercepts for 1-GMM and C2-GMM models. More specifically, the direction of raw bias was generally the same as the direction in intercepts change for both analysis models, particularly, when there was non-negligible raw bias. This finding is the same as from Wirth (2008) who found the directional correspondence between the non-invariance pattern and slope factor mean bias. With mean generating item loadings and intercepts being non-invariant over time, the model using an averaged scale score (i.e., 1-GMM) or constrained item scores (i.e., C2-GMM) essentially used inappropriate weights among items. With increased item parameters, for example, the weights were inappropriately set as 1, which resulted in underweighting items when using items aggregately. The uncounted non-invariance on the item parameters was consequently pushed into the growth rate estimates with the sign remained. The directional effect of item intercepts change on accuracy in slope factor mean recovery was not detected for F2-GMM when measurement invariance sustained on the referent item. Wirth (2008) also had the same finding where no bias was observed on the slope factor mean as long as generating item factor loadings and intercepts maintained partial invariance.

Given the above findings, under minimal L-MI violation and increased item intercepts, this study generally draws the same conclusion as Wirth (2008) in model comparison where F2-GMM with maintenance of partial measurement invariance was found robust in accurate recovery of slope factor means for both latent classes under various deviation of L-MI assumption. Performance of C2-GMM was (marginally) well and 1-GMM was not acceptable with deviation of L-MI. Under decreased item intercepts, the magnitude of bias by 1-GMM was smaller than C2-GMM over all L-MI patterns with some of the bias being negligible.

As no previous studies evaluated impact from L-MI deviation on growth parameters recovery with multiple latent classes, the reason why 1-GMM outperformed C2-GMM under simultaneously decreased item loadings and intercepts is hypothesized to be from the altered latent classes distance by L-MI deviation. Bias on the growth rate estimates from 1-GMM is not only influenced by the direction of non-invariant items, but also the latent class distance or overlap under the generating L-MI patterns. With simultaneously decreased item loadings and intercepts, the distance between the latent classes should remain the same as measurement invariance was simulated between the two classes. As a consequence, compared to minimal violation of L-MI, under decreased item intercepts, the bias on the slope factor mean was similar with magnitude under .10 for 1-GMM (mostly around .05). When there was competing effects between change on item intercepts (positive) and item factor loadings (negative) in their direction, drawing upon the same rationale, distances between estimated latent classes was smaller. Thus, it was harder for 1-GMM to distinguish the latent classes, which resulted in less accuracy in estimating the growth rates under the same generating L-MI with the opposite direction on the change in the non-invariant item intercepts and loadings.

4.2.1.2 RMSE

Figure 5 compares three analysis models on the slope factor mean RMSE under combinations of generating L-MI pattern and item intercepts change, separately for C1 and C2. The levels on correlated ME were averaged out as their RMSE values were similar. The order of performance among the analysis models in terms of RMSE resembled that of bias. F2-GMM was able to estimate precisely the slope factor means for both latent classes under all but NiLpT pattern. 1-GMM had larger RMSE than C2-GMM when item intercepts increased, but smaller RMSE when item intercepts decreased. Hence, the factors that impacted the bias also affected precision of slope factor mean estimates in a similar way.

4.2.2 Slope factor variance

4.2.2.1 Relative bias

ANOVA results (Table 11) suggested that only analysis model had more than small effect size in predicting relative bias on slope variance estimate in C2 while contamination level and L-MI generating pattern were significant in C1 besides analysis model. Moreover, interaction between analysis model with contamination % and MI pattern were also significant. *Figure 6* presented the relative bias on slope factor variance estimate by L-MI pattern and analysis model for both classes for side by side comparison.

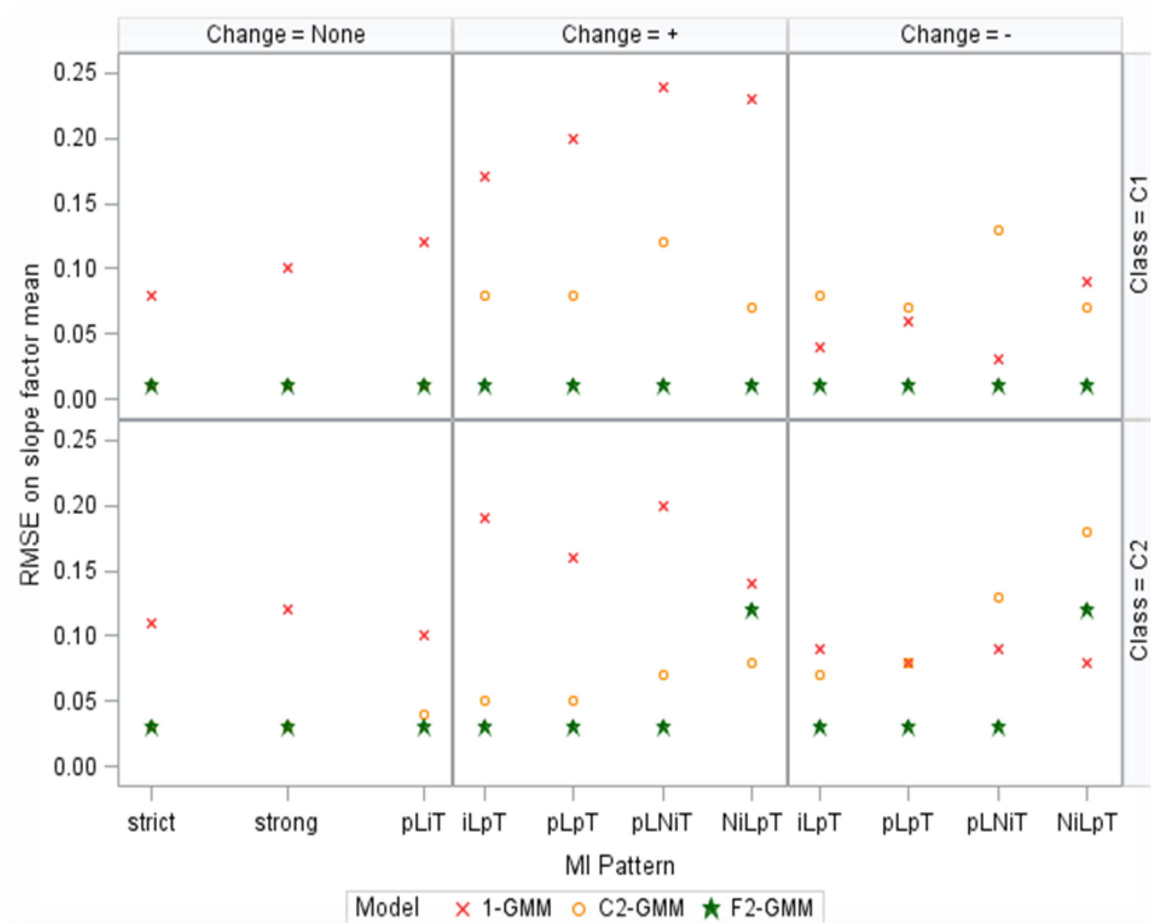

Figure 5. RMSE on class-specific slope factor mean under combinations of MI pattern and item intercepts change for each analysis model when MD = 5.

Figure 6 showed that 1-GMM had unacceptable relative bias in slope variance estimate. Slope factor variance was consistently overestimated by using 1-GMM. It was the case for both latent classes. F2-GMM performed better than C2-GMM where relative bias was acceptable among all but the NiLpT generating L-MI condition. On the other hand, C2-GMM produced negligible relative bias on slope factor variance with none or minimal violation of L-MI assumption (i.e., strict, strong, pLiT, and iLpT). With increased deviation of L-MI assumption (i.e., pLpT and pLNiT), the slope factor variance estimates for both latent classes became biased using C2-GMM, which was not able to estimate slope factor variance accurately just as F2-GMM under NiLpT condition.

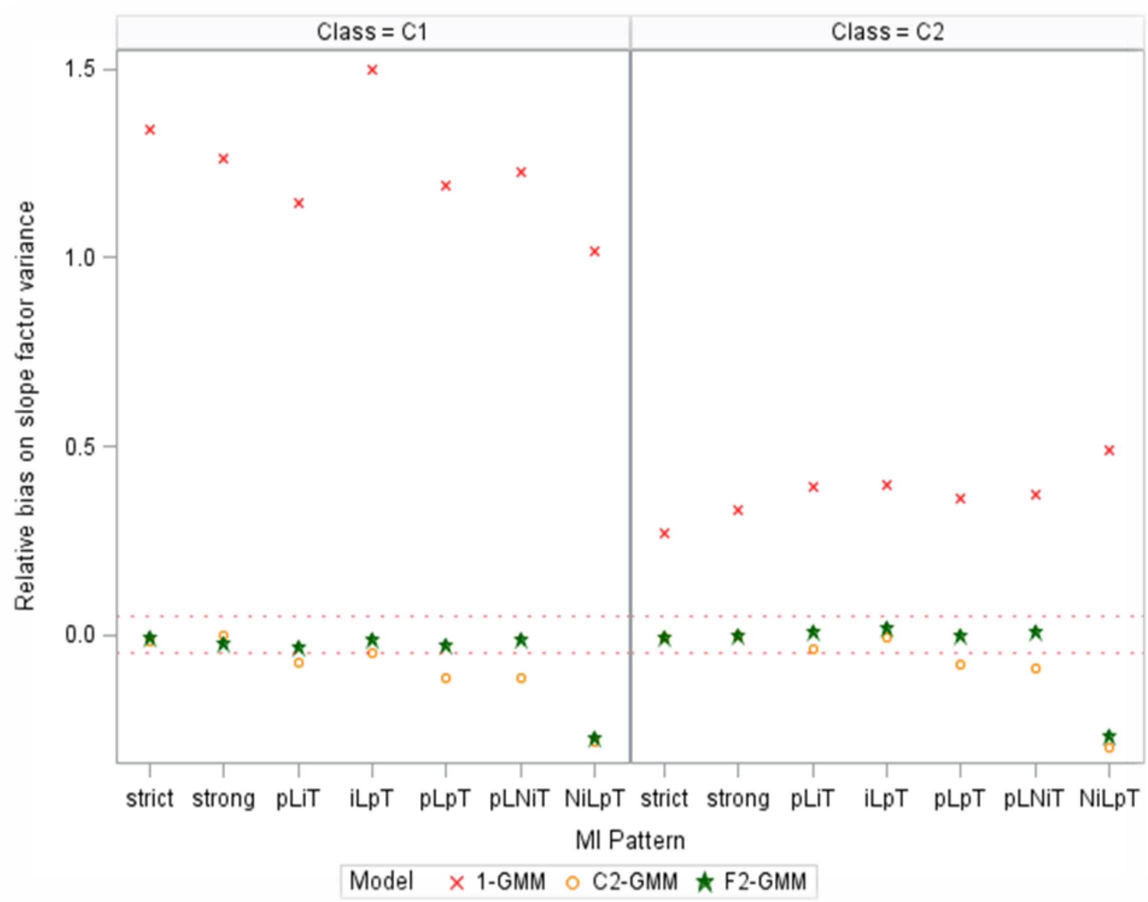

Figure 6. Relative bias on slope factor variance under generating MI patterns for each analysis model when MD = 5. Line of |.05| was drawn as reference.

The findings on the order among analysis models in preserving the bias in growth rates variance estimates are consistent as from Wirth (2008). In producing accurate growth rate variance estimates, the model in his study that constrained single item parameters performed the best compared to the mode that placed constraints on all within-item item parameters. The mean score model performed the worst in both studies with moderate to severe bias on the slope factor variance estimates.

4.2.2.2 RMSE

RMSEs on the slope factor variance estimates were consistently low and stable among generating measurement characteristics and the analysis models (Table 12). The variability on the parameter estimate was negligible for both latent classes. This might be due to the fact that the latent classes were separated extremely well and the within-class variability was relatively small. Under this premise, even with violation of L-MI, the models did not have a problem producing the estimate precisely.

Table 12. RMSE on slope factor variance by each of the analysis model for each latent class.

Model	C1	C2
1-GMM	0.01	0.03
C2-GMM	0.01	0.01
F2-GMM	0.01	0.01

4.2.3 Type I error and power

Empirical Type I error by each analysis model was calculated as the percent of replications incorrectly rejecting the true null slope factor mean in C1. Power of each analysis model was calculated as the percent of replications successfully rejecting the false null hypothesis that the slope factor mean was 0 in C2. When MD = 5, 1-GMM had inflated Type I error (i.e., .75) while the values from the 2nd-order models (i.e., .04 for C2-GMM and .03 for F2-GMM) were close to the nominal value. Power rates were plotted in *Figure 7* under all generating conditions, respectively, for each analysis model.

With the correlated ME, F2-GMM had greater than .80 power among various generating L-MI patterns. The only exception was in NiLpT condition where there was almost no power by F2-GMM. The use of 1-GMM resulted with acceptable power when there was no or minimal violation of L-MI assumption (strict, strong, pLiT). When item intercepts decreased with simultaneous decrease in item factor loadings, 1-GMM still had acceptable power rates under varying generating L-MI patterns, which can be due to the remained generating latent classes distance after L-MI violation. When there was competition on the directional change between item intercepts and item factor loadings (i.e., increased item intercepts with decreased item factor loadings), the ability for 1-GMM correctly rejecting the null hypothesis was largely reduced to the extent that the power was no longer acceptable. Similar to the hypothesized reason for the poor performance of 1-GMM in estimating slope factor means under increased item intercepts, 1-GMM's poor performance in power can be the result of more overlapped latent classes (i.e., smaller between-class distance) after L-MI deviation. As for C2-GMM, it only had high enough power when item intercepts remained invariant (i.e., strict, strong, pLiT). Regardless of the direction on the item intercepts change, C2-GMM was not able to reject the null hypothesis with a satisfactory rate.

Generally, the pattern of power rates for each analysis model was similar to their counterpart, compared to the scenario where no correlated ME was simulated, as far as using .80 as cut-off value in drawing conclusions on the performance for the analysis models. The exception was in 1-GMM under minimal MI violation, where an over-fitted measurement model reduced the power to slightly below .80.

4.2.4 Classification

Numbers of individuals assigned to C1 and C2 were averaged at 800 and 200 without much variability (SD = 9) among the simulation cells under MD = 5. Even though the overall number of individuals classified into each latent class was basically correct, the individuals were not necessarily classified into their generating latent class. Partially nested ANOVAs suggested that only analysis model significantly (eta-square = .02) impacted the individual classification accuracy for both latent classes. In general, C2-GMM had consistently the highest accuracy in individual classification rate but the classification by any of the analysis model was not satisfactory. More specifically, the classification accuracy aggregated between both latent classes were 59%, 73%, 57% for 1-GMM, C2-GMM, and F2-GMM. Hence, about 590 individuals were correctly classified into the latent class that he/she was generated from using 1-GMM, and so forth for C2-GMM and F2-GMM which respectively classified accurately about 730 and 570 individuals over the two latent classes.

The recovery of the mixture proportion and individual classification found under MD =5 in this study partially agreed with the findings from Liu (2012) who studied both raw count and accurate count in 1-GMM with varying between-class distances. He found that the unbalanced mixture proportion was not recovered well and this does not depend on the between-class distance. It is not the case in this study where the mixture proportions in the unit of sample were recovered well, and the possible reason is that the latent classes were extremely well separated, when compared to the maximum MD = 1.6 in Liu's study. On the other hand, classification accuracy in his study under the most optimal between-class distance was found to be similar as the numbers in this study. Hence, with extremely well separated latent classes, the mixture proportions could be well recovered but not classification accuracy at the individual level.

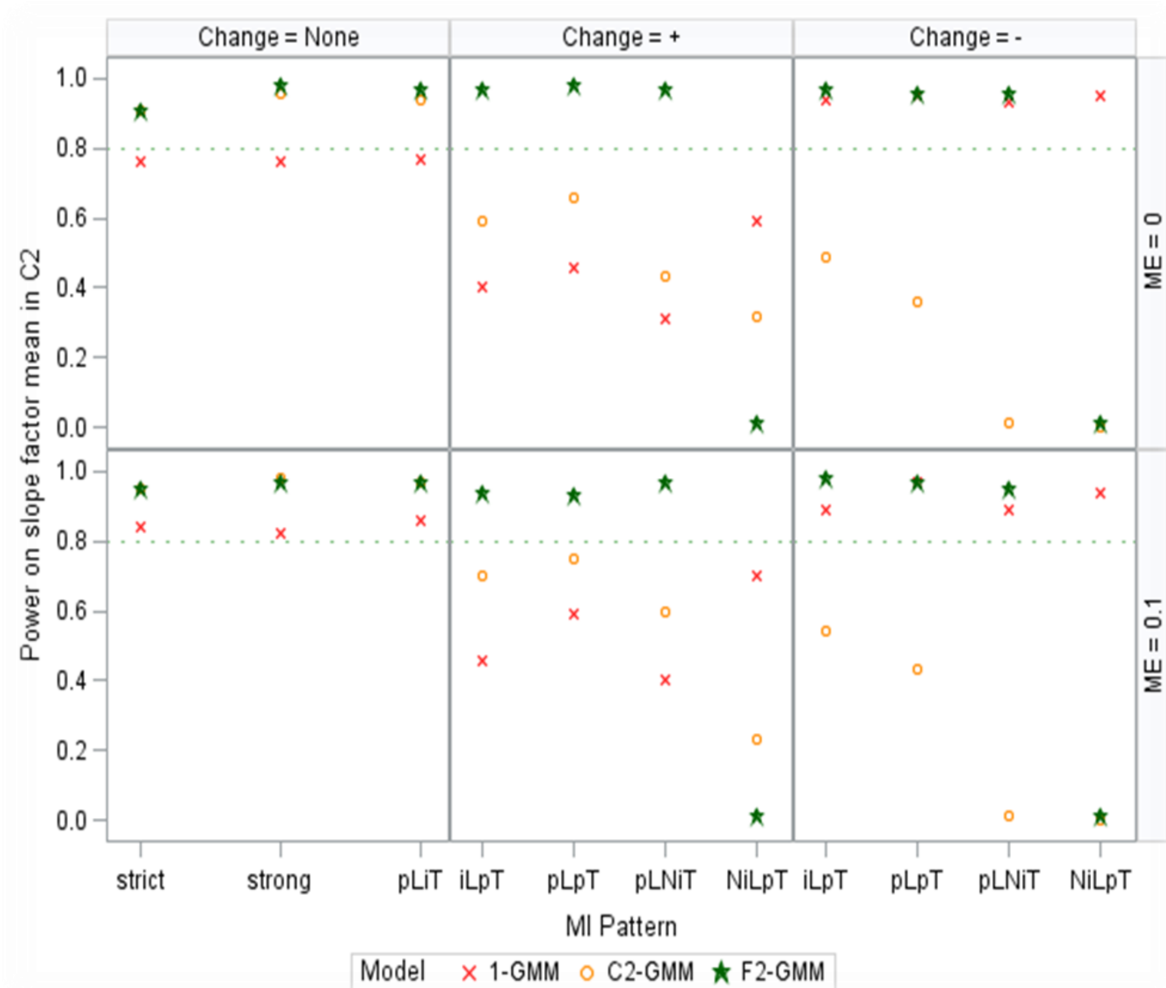

Figure 7. Empirical power rates detecting other-than-zero slope factor mean under varying conditions when MD = 5. ME = Measurement error correlation. Line of .80 was drawn as reference.

4.3 GROWTH CHARACTERISTICS RECOVERY UNDER MD = 1.5

The ANOVA results were summarized in Table 13 for MD = 1.5. The table had the same set up compared to the ANOVA table for growth rate estimates under MD = 5 with the same omitted effects and symbols indicating the nesting structure.

Analysis model was a significant predictor for almost all growth rate estimates with effect sizes ranging from .04 to .21. The only exception is that the accuracy in the slope factor mean recovery for the dominant class was not explained by the main effect of analysis model, but by the interaction between the analysis model and item intercepts change. Item intercepts change factor was practical significant in recovery of slope factor mean estimates for both latent classes.

4.3.1 Slope factor mean

4.3.1.1 Raw bias

Based on the ANOVA results, main effects of item intercepts change and analysis model significantly impacted the raw bias of the slope factor means for the rare class (C2). Their interaction also had explained significant proportion of raw biases in the dominant class (C1). Mean raw bias under each level of item intercepts change for each analysis model was plotted in *Figure 8*, for C1 and C2.

When item intercepts were invariant (i.e., strict, strong, and pLiT), there was negligible raw bias by F2-GMM for both latent classes. Under the same generating L-MI pattern, C2-GMM produced generally unbiased slope factor means, with no bias in C2 and marginally acceptable bias in C1. As for 1-GMM, even with minimal violation of L-MI assumption, slope factor means were estimated with much less accuracy.

With increased item intercepts, slope factor mean in C1 was estimated with no bias regardless of the analysis model used. The slope factor mean in C2, however, was (marginally) accurately estimated by the 2nd-order models but not the 1-GMM where raw bias was unacceptably large. With decreased item intercepts, slope factor mean was only estimated accurately by F2-GMM.

Table 13. ANOVA results investigating class-specific growth parameter estimates accuracy when MD = 1.5.

Latent Class	Parameter Estimate	Main Effects: Model	Main Effects: Contamination	Main Effects: MI Pattern	Main Effects: Intercepts Change	2-way Interactions: Model × Contamination %	2-way Interactions: Model × MI Pattern	2-way Interactions: Model × Intercepts Change
C1	Slope Mean*	-----	-----	-----	0.04	-----	-----	0.02
	Slope Variance	0.04	-----	-----	-----	-----	-----	-----
C2	Slope Mean*	0.21	-----	-----	0.04	-----	-----	-----
	Slope Variance	0.07	-----	-----	-----	-----	-----	-----

Note: * indicated raw bias was used on slope factor mean as outcome in the ANOVA instead of relative bias.

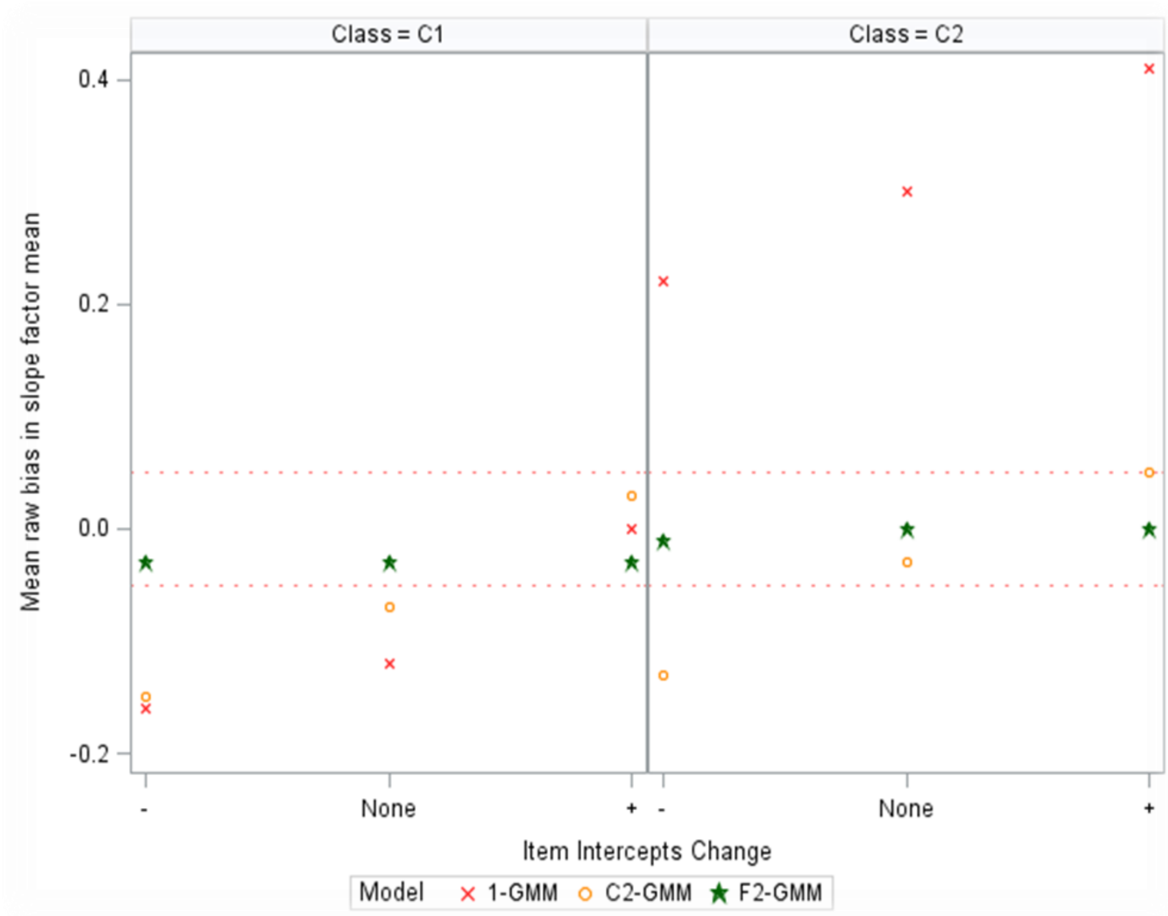

Figure 8. Mean raw bias on class-specific slope factor mean under combinations of MI pattern and item intercepts change for each analysis model when MD = 1.5. Line of |.05| was drawn as reference.

Hence, under MD = 1.5, the bias magnitude was acceptable using F2-GMM, regardless of item intercepts change and which latent class the slope factor mean parameter was from. The magnitude of bias produced by C2-GMM was similar between the latent classes under the same item intercepts change condition. For 1-GMM, slope factor mean was less accurately estimated in C2 than in C1, reflected by the upward shift of mean raw bias under the same item intercepts change condition.

4.3.1.2 RMSE

RMSEs on the slope factor mean estimates under MD = 1.5 for the 2nd-order models were similar between levels of correlated ME while this is not the case for 1-GMM. In general, when there was minimal violation of measurement invariance, slope factor mean was estimated with more variability (i.e., larger RMSE) with presence of within-item unique factor correlation than when there was an absence in correlated ME. The conclusion applied to the dominant class but not the rare class. The difference between the latent classes was tangled with generating MI pattern and direction in item intercepts change where no consistent and distinctive pattern on RMSE was found for 1-GMM. Nevertheless, all of these RMSEs by 1-GMM were much larger than the ones produced by the 2nd-order models under the same condition. To enhance interpretability compared to MD = 5, RMSEs were averaged out between absence and presence of correlated ME and the values were plotted in *Figure 9*.

Slope factor mean estimates were more variable for the rare class than for the dominant class. Regardless of the generating L-MI characteristics (i.e., contamination %, MI pattern, and item intercepts change), F2-GMM was able to produce stable slope factor mean estimate. The variability in the slope factor mean was slightly larger in C2 than in C1 but both variability were close to 0. C2-GMM produced larger variability in the estimate than F2-GMM under any generating condition. The difference was smaller in C2 but more obvious in C1. 1-GMM, on the other hand, had consistently larger variability in estimating the slope factor mean parameter than the 2-nd order models.

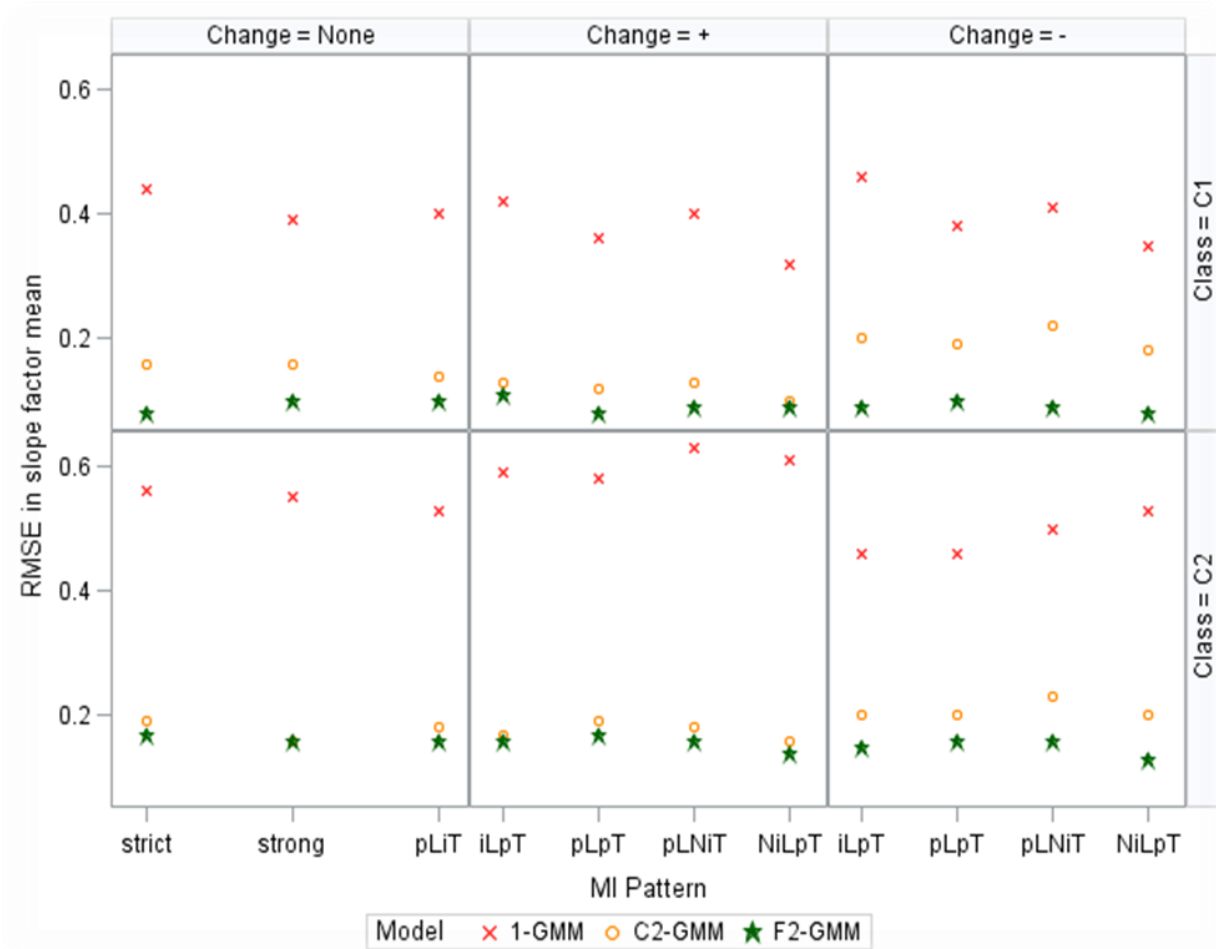

Figure 9. RMSE on slope factor mean under generating MI patterns for each analysis model when MD = 1.5.

Combined with the results on raw bias, F2-GMM was able to produce both accurate and precise slope factor mean estimates, especially for the dominant class under MD = 1.5. C2-GMM performed as the second best in recovering the slope factor means. The parameters were estimated with (marginal) accuracy and precision when there was no or increased item intercepts change. The model's performance got worse when all of the item parameters had non-invariance in the same direction. When both item intercepts and item factor loadings decreased, the slope factor means were not estimated accurately meanwhile with considerable variability. 1-GMM retuned slope factor mean estimates with large variability (i.e., at least 2 times larger than the 2-

nd order models). Even under none or minimal violation of L-MI assumption (i.e., strict, strong, pLiT), there was large variability in the slope factor mean estimates by 1-GMM, which was considered to be caused by the large deviation of the true model which might be made worse with less indistinguishable latent classes.

4.3.2 Slope factor variance

4.3.2.1 Relative bias

Accuracy in slope factor variance estimates were affected by the analysis model. 1-GMM produced large positive relative bias on the slope factor variance estimates for both latent classes (Table 14). Both of the 2nd-order models were found to have underestimated slope factor variance estimates. Between the models, F2-GMM was able to return unbiased slope factor variance estimates, with negligible relative bias in C2 and marginally acceptable in C1. C2-GMM also had bias on the slope factor variance estimate under control for C2 and marginally for C1.

Table 14. Relative bias in the slope factor variance by analysis model for each latent class under MD = 1.5.

Model	C1	C2
1-GMM	2.2	3.94
C2-GMM	-0.12	-0.05
F2-GMM	-0.08	-0.01

4.3.2.2 RMSE

Conclusion on the magnitude in the RMSEs on the slope factor variance were similar to the conclusion on the relative bias, for each of the analysis model. 1-GMM, regardless of the generating MI conditions, was not able to estimate the parameter precisely with the disappointing performance more obvious in the rare class (*Figure 10*).

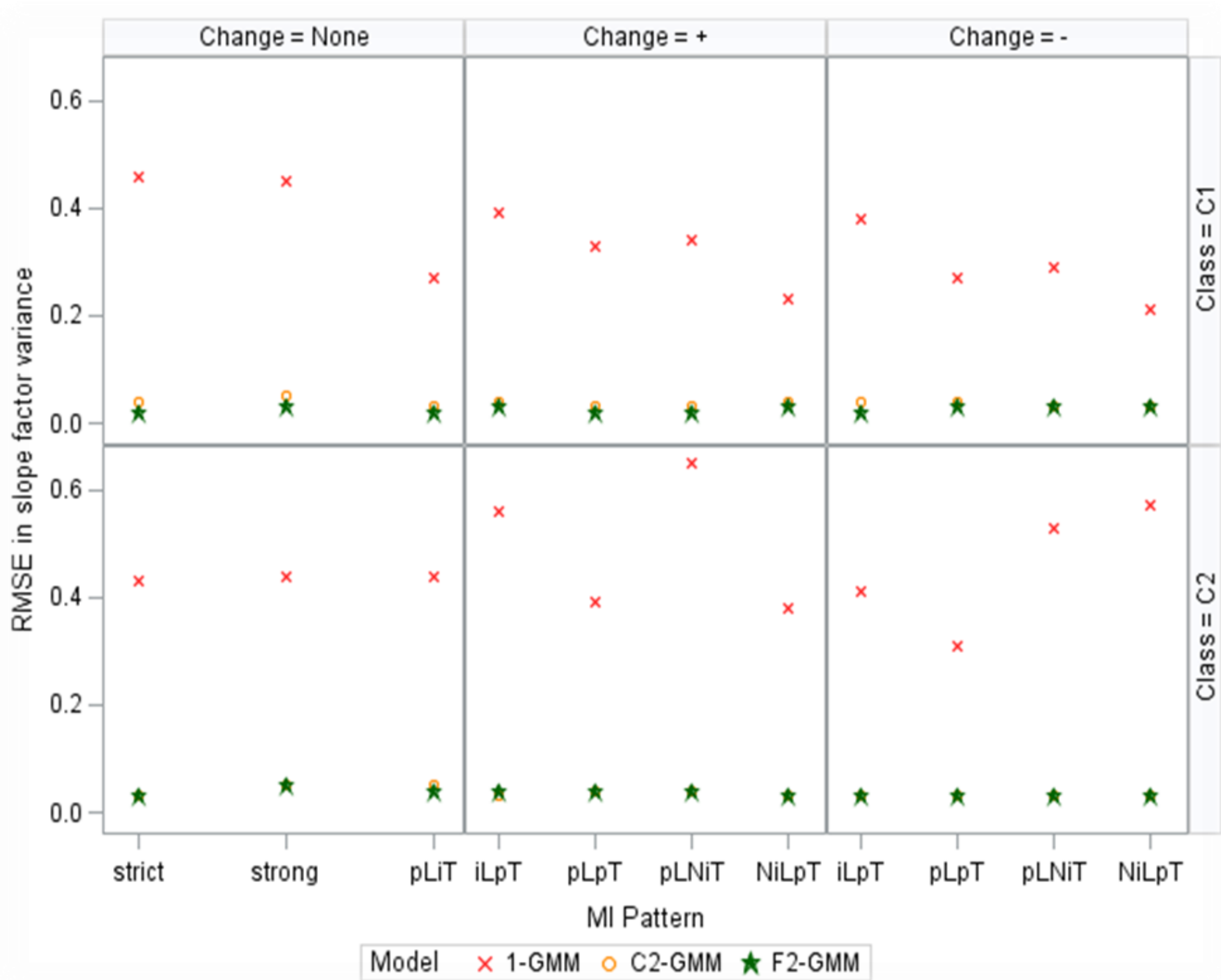

Figure 10. RMSE on slope factor variance under generating MI patterns for each analysis model when MD = 1.5.

4.3.3 Type I error and power

Type I error and power were calculated in the same way as described when MD = 5. The increased overlap between the two latent classes highly impacted the performance of the 2nd-order models in both Type I error and power. More specifically, Type I error rates were .45, .29 for C2-GMM and F2-GMM, respectively while the rate remained unacceptably high for 1-GMM (.74). Moreover, none of the analysis model was able to achieve .80 for power rate under any of the generating measurement characteristics (*Figure 11*). Hence, none of the analysis model was able to correctly retaining 1) the null hypothesis for the slope factor mean in the dominant class, and 2) the alternative hypothesis for the slope factor mean in the rare class when the classes were more overlapped.

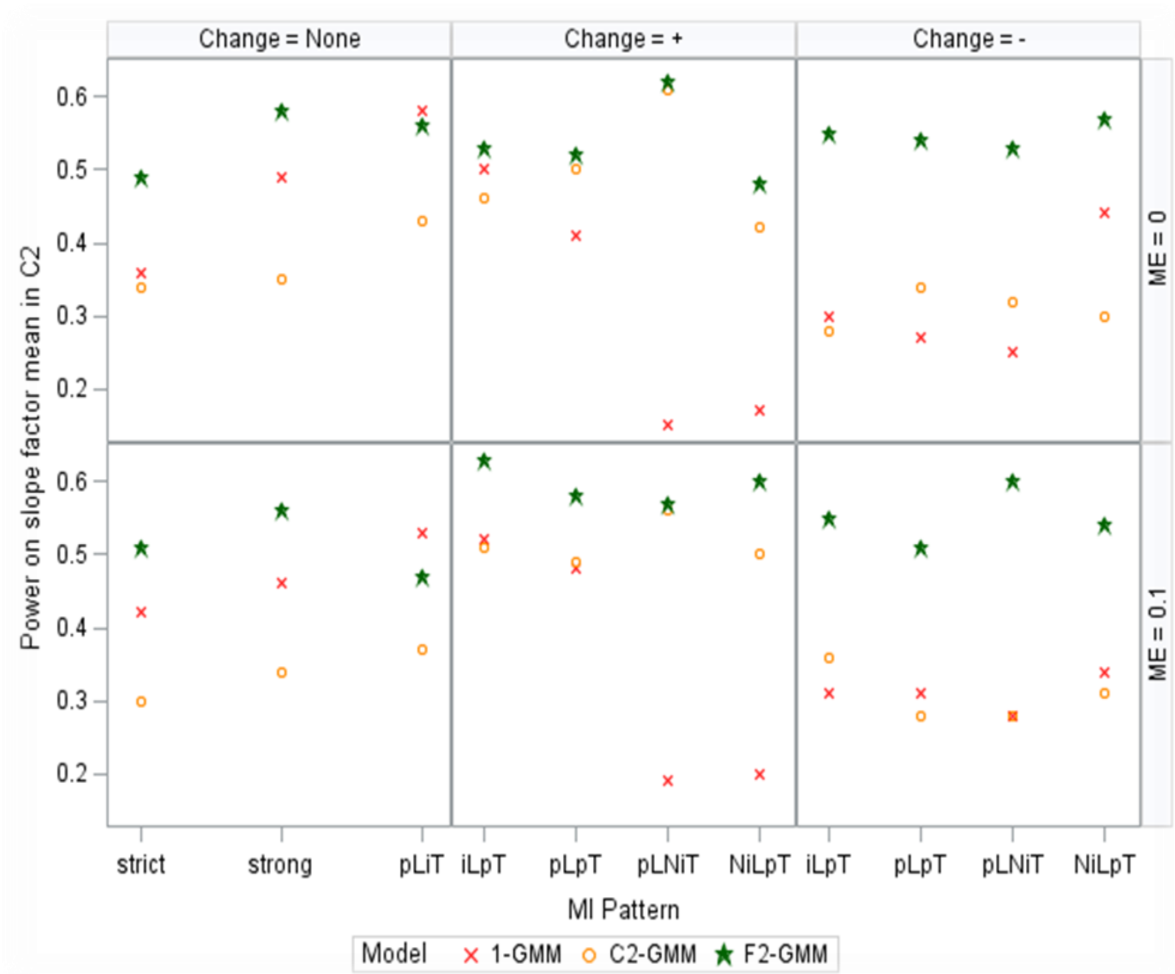

Figure 11. Empirical power rates detecting other-than-zero slope factor mean under varying conditions when MD = 1.5.

4.3.4 Classification

Analysis model impacted the recovery of the raw count in individual assignment into each of the latent classes with an eta-square of .03 for both latent classes. Estimated mixture proportions in unit of sample size for each latent class by each analysis model are summarized in Table 15. As can be seen, for the rare class, more than 200 (number of individuals generated from the rare class) individuals were estimated to be from the rare class. Meanwhile, the size of the dominant class was shrinking. It indicated that when latent classes were not separated as well, GMMs tended to decrease the difference in the sizes between the latent classes.

This finding was the same as Liu (2012)'s study where he found under less distinguishable latent classes, 1-GMM artificially increased mixture proportion estimate for the rare class while decreased the estimate for the dominant class. With smaller latent class distances (MD = 1.5 in the current study and MD = 1.6 in Liu's study), both studies found that mixture proportion recovery was not satisfactory. The accuracy in the classification by all the analysis models remained not satisfactory. Among the analysis models, C2-GMM was found with the most severe loss in accuracy (43% compared to 73%). The other two models had worse performance but not as much with 55% for 1-GMM as compared to 59%, 52% for F2-GMM as compared to 57%. Hence, 550, 430, and 520 individuals were correctly classified into their generating latent class by using 1-GMM, C2-GMM, and F2-GMM.

Table 15. Raw count by analysis model for each latent class

Model	C1 Count	C2 Count
1-GMM	537	436
C2-GMM	494	506
F2-GMM	663	337

4.4 SUMMARY OF RESULTS

This section summarizes the results by the design factors in the order of 1) analysis model, 2) correlated measurement error, 3) longitudinal measurement non-invariance characteristics, and 4) latent class distance.

4.4.1 Analysis model

The three analysis models were compared in terms of the recovery of the growth characteristics under each between-class distance in Table 16 and Table 17. In general, F2-GMM with true identification invariance is the most robust model under various deviation of L-MI deviation in recovery of slope factor estimates. The comparative performance between C2-GMM and 1-GMM was not universal over different L-MI generating configurations. Rather, the order of these two models' performance is mixed depending on the contamination level, MI pattern, item intercepts change, and between-class distance.

4.4.2 Correlated ME

Difference between over- and appropriately-specified measurement models with respect to within-item correlated ME was not systematically associated with recovery of growth rates. This finding was somewhat different from the findings in Wirth (2008)'s study who found ignored presence of the small amount of within-item unique factor correlation in a two-step procedure impacted recovery of growth factor variance estimates. The difference highlights that when common factor scores and growth factor scores are modeled sequentially in two steps, even when the within-item unique factor correlation is small, the model used to estimate common factor scores need to accommodate the with-item relationship. Otherwise, the growth factor variance estimates were impacted by inappropriately estimated common factor scores. However, when common factor scores and growth factor scores are modeled simultaneously, the measurement model with over-specified within-item unique factor correlation did not hurt the accuracy in the underlying growth rates recovery with the magnitude of correlation being small.

4.4.3 Contamination level, L-MI pattern and non-invariant item intercepts directional change

Item intercepts change was nested within MI pattern which was nested within contamination level in the design. The three factors had effect on almost all aspects in growth rates recovery except precision in class-specific slope factor variance under MD = 5. The worsen recovery of the growth parameters under various measurement conditions were tangled with a couple of other factors including 1) mechanism of the scoring from a particular analysis model, 2) specific growth parameter, 3) specific latent class, and 4) competing directional change on the item parameters. Hence, no universal conclusion on the impact of the three factors can be made. Rather, the interaction effects between these factors should be looked at. The three factors influenced the magnitude of deviation on growth rates recovery under MD = 5 where it can be considered as the influence from deviated L-MI on class-specific growth parameter recovery without influence from between-class distance. The performance on the recovery of the parameters did not necessarily get worse with more L-MI deviation. However, when the generating latent classes were closer (MD = 1.5), the effects of these three factors were no longer Significant except directional change on non-invariant item intercepts.

Table 16. Summary table of recovery in growth characteristics by each analysis model when MD = 5.

Outcome of interest	Accuracy			Direction of Bias			Precision		
	F2-GMM	C2-GMM	1-GMM	F2-GMM	C2-GMM	1-GMM	F2-GMM	C2-GMM	1-GMM
Slope Mean (C1)	Yes	No[2]	No[2]	------	Same as item intercepts	Same as item intercepts	Yes	No[2]	No[2]
Slope Variance (C1)	Yes[1]	No[2]	No	------	Negative	Positive	Yes	Yes	Yes
Type I error on Slope Mean (C1)	Yes	Yes	No						
Slope Mean (C2)	Yes[1]	No[2]	No[2]	Negative	Same as item intercepts	Same as item intercepts	Yes[1]	No[2]	No
Slope Variance (C2)	Yes[1]	No[2]	No	Negative	Same as item intercepts	Positive	Yes	Yes	Yes
Power on Slope Mean (C2)	Yes[1]	No[2]	No[2]						
Raw Count	Yes	Yes	Yes	------	------	------			
Overall Accuracy	No	No	No						

Note: 1 indicated exception under *NiLpT* condition. 2 meant dependent on the generating measurement characteristics including contamination %, MI pattern, and item intercepts change. 3 meant marginally.

Table 17. Summary table of recovery in growth characteristics by each analysis model when MD = 1.5

Outcome of interest	Accuracy			Direction of Bias			Precision		
	F2-GMM	C2-GMM	1-GMM	F2-GMM	C2-GMM	1-GMM	F2-GMM	C2-GMM	1-GMM
Slope Mean (C1)	Yes	No[2]	No[2]	------	Same as item intercepts	Opposite between classes	Yes	No[2]	No
Slope Variance (C1)	Yes[3]	No	No	------	Negative	Positive	Yes	Yes	No
Type I error on Slope Mean (C1)	No	No	No						
Slope Mean (C2)	Yes	No[2]	No	------	Same as item intercepts	Opposite between classes	Yes[3]	Yes[3]	No
Slope Variance (C2)	Yes	Yes	No	------	------	Positive	Yes	Yes	No
Power on Slope Mean (C2)	No	No	No						
Raw Count	No	No	No						
Overall Accuracy	No	No	No						

Note: 1 indicated exception under *NiLpT* condition. 2 meant dependent on the generating item intercepts change. 3 meant marginally.

Item intercepts change or the competing directional change on item intercepts and item factor loadings was found to impact the direction in bias of slope factor mean estimates, for the analysis models using aggregated relation among item scores (i.e., 1-GMM). More specifically, the direction in the change from the item parameters was generally consistent as the direction in the bias on the slope factor mean. The results intuitively made sense as the location of scale score at each time was defined by both the magnitude and the sign of the item parameters. Hence, when there was decreased non-invariance patterns on both of the item intercepts and item factor loadings, the model-implied scale score at each time decreased. To capture the decrease, the growth rate estimate would decrease, and vice versa, when with an increased non-invariance pattern.

One thing to note was the effect from the pattern of NiLpT. Among the varying measurement characteristics, NiLpT condition mimicked the most deviated L-MI assumption. All item factor loadings and 3 item intercepts were decreasing under this condition. When latent classes were separated well (to make inference without influence from increased overlap), the pattern was found associated with unacceptable recovery on both fixed and random effects for slope factor in the rare class. Moreover, it produced nearly no power in detecting the slope factor mean in the rare class.

4.4.4 Latent class distance

The different latent class distances presented different conclusions on effect of L-MI violation, model robustness comparison, Type I error and power on slope factor mean recovery and necessity of identification invariance assumption.

Systematic effect of L-MI violation in predicting growth rates recovery was washed out when latent classes were less distinctive. The only exception is the directional change in the non-invariant item intercepts which significantly impacted the slope factor mean recovery. It indicated that the generating between-class distance is interacted with the occurrence of L-MI violation. Order of performance between C2-GMM and 1-GMM was also dependent on generating latent class distance and specific directional change on the non-invariant item parameters. Type I error was controlled and power was satisfactory by the most robust GMM model (i.e., F2-GMM) under maintenance of identification invariance when classes were separately extremely well. With increased overlap between the latent classes, the Type I error was inflated and power was low regardless of the analysis model. As for the identification invariance assumption, it was critical in recovery of slope factor estimates under extremely distinct classes. However, when the latent classes were more clustered, whether the single item used for constraining item parameters truly met identification invariance assumption did not matter anymore. With the violation of identification invariance, bias on the slope factor estimates produced by F2-GMM was under control.

5.0 DISCUSSION

GMM has been used to study heterogeneous developmental pathways for various subpopulations in different disciplines where class-specific latent growth trajectory has been extracted with clinically meaningful interpretations. To draw valid conclusions on longitudinal change in the measured construct, researchers need to make sure the items repeatedly administrated have similar meaning for the respondents over time. Otherwise, comparing mean differences across occasions in order to reach conclusion on patterns of change is nothing different from comparing apples to oranges.

From previous research, deviation from L-MI assumption adversely impacted the growth characteristics which could be estimated with lack of accuracy and precision. Meanwhile, the conclusion on the shape of the growth trajectory could also be altered. As GMM models several growth trajectories, L-MI is projected to influence growth recovery when multiple latent classes exist. To our knowledge, this study is the first one that examined the impact from L-MI assumption deviation on growth characteristics recovery in GMM framework. Recovery on class-specific fixed and random effects on the slope factors (slope factor means and slope factor variances) and marginal and individual classification rates were investigated using a Monte Carlo simulation. Six factors were systematically manipulated in studying the impact of L-MI assumption violation and robustness of three alternative GMMs: directional change in non-invariant item intercepts, patterns of item loadings and item intercepts, percent of items containing a set of non-invariant item parameters, presence of time-adjacent within-item correlated measurement error, latent class distances, and different GMM analysis models. Accuracy, precision, Type I error, and power were examined on the slope factor parameter estimates. Additionally, mixture proportion and individual classification were assessed.

This chapter summarizes and discusses the results in the order of the general research questions presented in Chapter 3, followed by a discussion of the limitation of the study, direction for future research, and implication based on the results for applied researchers.

To recall, the general research questions this study aimed to answer are:

1) Are growth factors in GMM affected when longitudinal measurement non-invariance is present? If yes, what factors contribute to the biased/imprecise growth parameters?
2) Does the L-MI deviation impact the classification accuracy and if yes, what factors contribute to poor classification accuracy?
3) Do the factors affect the growth parameters and classification accuracy in the same way with the same magnitude for each latent class?

5.1 RQ1: IMPACT ON GROWTH FACTORS

Growth factor estimates recovery in multiple latent classes was found to be impacted by the violation of L-MI, when there was a discrepancy between the generating L-MI configurations and the analysis model used to obtain the growth factor estimates.

Longitudinal measurement non-invariance pattern. Increased bias and decreased precision on the slope factor estimates were found with higher contamination level and more deviated L-MI pattern. This finding was the same as in LGM with one latent class from Wirth

(2008) and Olivera-Aguilar (2013) where the slope factor mean and variability showed the largest degree of bias with violation of L-MI assumption. It indicated the bias on the growth rate estimate was related to not only the number of non-invariant items but also the non-invariant item parameters including both item factor loadings and intercepts.

Direction of item intercept change. The directional change in non-invariant item intercepts had an impact on the direction of bias in slope factor mean estimates for 1-GMM and C2-GMM. This study found the direction in the bias on slope factor mean estimates was the same as the directional change on the non-invariant item intercepts with decreased item factor loading and increased item intercepts, which was similarly concluded in Wirth (2008) and Olivera-Aguilar (2013) in one latent class. Hence, this study generalizes the finding into both latent classes with competing non-invariance direction between item factor loadings and intercepts for both 1-GMM and C2-GMM, regardless of generating between-class distance. This study additionally found the opposite direction in class-specific slope factor mean estimate bias for 1-GMM under same directional change on non-invariant item intercepts when between-class distance was smaller.

Within-item error correlation. In the current study, the factor of within-item measurement error correlation was not associated with explained variability in the accuracy or precision of the growth rate estimates (both means and variances). The conclusion was different from Wirth (2008)'s study where the measurement error correlation, with the same magnitude, was practically significant in impacting the accuracy in growth factor variability. The difference in the conclusion might be due to the fact that factor scores (both 1st- and 2nd-order) were estimated simultaneously under a specified measurement model in the current study, while Wirth used a two-stage process to estimate the growth factors using calculated latent scores. Even though the measurement model with respect to within-item unique factor correlation was over-specified, the influence of the mis-specification on the recovery of slope factor variability estimates in a concurrent step (as in the current study) was not as strong as the systematically generated L-MI configuration combined with the alternative scoring models. In contrast, when the growth factor scores were estimated in a subsequent step using the factor scores obtained first (as in Wirth's study), the analysis model's ability in correctly reproducing the within-item relationship was more important on recovery of the growth factor scores.

Analysis model. Among the three analysis models, F2-GMM outperformed C2-GMM and the first-order model in terms of their robustness in obtaining accurate growth factor estimates. Wirth (2008) made the same conclusion in LGM where the model that assumed L-MI on one item was the most robust followed by the same model but with assumed L-MI on all items. The model based on mean score composite was the least robust in recovering growth factor estimates once there was deviation of L-MI assumption.

Function of single constrained item for identification invariance. When identification invariance held, F2-GMM was able to accommodate the non-invariant item parameters so that the underlying (common and growth) factor parameters were recovered well. When identification invariance did not hold, it resulted in unacceptable large bias on the average growth rate estimate using the F2-GMM which inappropriately put constraint on non-invariant item factor loadings. These findings corroborated what were found in Wirth (2008) on the impact of identification invariance in LGM. However, it was not found previously that the dominant class was free of the impact from violation of L-MI and wrong identification invariance on the estimated growth rate.

Latent class distance. The impact of L-MI deviation was dependent on how well latent classes are separated. The interacted relationship was reflected on growth factor recovery, bias direction on growth rates estimates, and the functionality of identification invariance. When the latent classes were separated extremely well, L-MI violation had systematic impact on the recovery of growth rate estimates. When they were less distinguishable, the original distance between the latent classes was confounded with the changing measurement characteristics so that effects from contamination level and generating L-MI patterns were washed out leaving only choice of analysis model being most influential on recovery of the growth rate estimates followed by non-invariant item intercepts change direction.

The study additionally found the modified conclusion on Type I error and power on the slope factor mean estimates. When between-class distance was really large, Type I error and power were impacted by violation of L-MI assumption but the impact was reduced to negligibility when the model used to obtain the growth factor estimates corresponded to the true generating model (i.e., with identification invariance maintained) under various L-MI configurations. The robustness of F2-GMM disappeared with increased overlap between latent classes in addition to the changing measurement characteristics where Type I error was not acceptable and power was not satisfactory. Despite the difference in the above conclusions, the shrunk distance between the latent classes had the following similar patterns as compared to the situation where between-class separation was extremely well. Unsatisfactory performance in classification was still only related to analysis model. Over-fitted within-item unique factor correlation similarly did not impact accuracy and precision on growth rate estimates. Order of the performance among the analysis models preserved with F2-GMM being the most robust under varying L-MI configurations followed by C2-GMM. 1-GMM was not able to recover growth rate estimates well. The hierarchy among the models was reflected from bias and precision on both the growth factor means and variances.

The conclusion on directional bias on the slope factor mean estimates from non-invariant item intercepts were similar within each latent class for C2-GMM and 1-GMM where shifted up bias was observed with increased item intercepts as compared to invariant item intercepts, and vice versa with decreased item intercepts. However, with decreased between-class distance, the direction of bias on the slope factor mean estimates by 1-GMM were found to be opposite between the two latent classes under the same item intercepts change level. It can be considered as the first-order model was trying to maximally separate the location of the growth rate estimates for the two classes when they were not as easily distinguishable. As for the identification invariance, presence/absence of it did not impact the growth rates recovery, unlike when classes were minimally overlapped. Moreover, the failure of maintaining identification invariance did not result in differences on class-specific growth rate recovery.

5.2 RQ2: IMPACT ON CLASSIFICATION RATE

The recovery on the marginal classification (i.e., mixture proportion in the unit of sample size) and accurate overall classification among individuals from the latent classes were not influenced by L-MI deviation. Marginal classification was recovered well when latent classes were separated really well. Despite this fact, accuracy in the individual classification was not satisfactory with L-MI maintenance or not using any of the analysis model. This finding was somewhat counter-intuitive. The reason is that individuals are classified into respective latent classes based on two factors. One is the identified location of the growth factor (its mean) and the other is the estimated individual deviation from the growth factor means so that each individual has an estimated growth factor score (e.g., individual slope). If the violation in L-MI has been appropriately accounted for where the growth factor means are recovered well, performance on the individual estimates on the growth factor scores are expected to be reasonably well so that individual classification should be decent. However, it was not the case even under the most stringent L-MI assumption (i.e., strict invariance).

In Mplus, during classification, each individual was assigned into the latent class with a higher estimated posterior probability. A post-hoc analysis was performed on 4 replications under MD = 5 to check the differences in the estimated posterior probabilities between the latent classes. The descriptive statistics of the posterior probability for each generating latent class into the respective correct and incorrect latent class were summarized in Table 18 by each of the analysis model. Each mean represents the average posterior probability among the individuals classified into the predicted latent class. Each standard deviation indicates how spread out the estimated posterior probabilities are. Minimum and maximum columns show the lower and upper bound of the estimated posterior probabilities under the specific predicted class.

From the table, it can be seen that, on average, the average estimated posterior probabilities for the correct classification (i.e., predicted class being the same as the generating class) were much higher than the average posterior probabilities for the incorrect classification (i.e., predicted class being different from the generating class). It was the case for all analysis models regardless of the generating latent class. Additionally, there was no overlap on the ranges of the estimated posterior probabilities between the correct classification and the incorrect classification by each analysis model (e.g., .63-.99 for predicted C1 from C1 and .00-.39 for predicted C2 from C1 by 1-GMM). The conclusion applied to all analysis models among all conditions. Hence, the poor individual classifications cannot be explained by the differences in estimated posterior probabilities.

Table 18. Descriptive statistics of posterior probabilities for incorrect and correct classification by each analysis model

Generating Class	Predicted Class	Model	Mean	Std Dev	Minimum	Maximum
C1	C1	1-GMM	0.88	0.13	0.63	0.99
	C2		0.16	0.14	0.00	0.39
	C1	C2-GMM	0.94	0.10	0.66	1.00
	C2		0.19	0.11	0.01	0.36
	C1	F2-GMM	0.90	0.11	0.66	1.00
	C2		0.16	0.12	0.01	0.36
C2	C2	1-GMM	0.89	0.09	0.68	0.98
	C1		0.15	0.10	0.01	0.33
	C2	C2-GMM	0.92	0.09	0.66	0.99
	C1		0.18	0.09	0.02	0.31
	C2	F2-GMM	0.89	0.10	0.66	0.98
	C1		0.16	0.10	0.01	0.32

On the other hand, the results on the poor individual classification were generally consistent as the findings from Liu (2012) who concluded with poor classification accuracy with imbalanced mixture proportions. In his study, the percentage of correct assignment of class membership ranged in the sixty's to seventy's with minimal between-class overlap, which are similar to the numbers in this study under similar total sample size and sample size for respective latent classes.

5.3 RQ3: BETWEEN-CLASS DIFFERENCE IN THE IMPACT ON GROWTH FACTORS

Non-invariant item intercepts change direction and analysis model were found to impact slope factor mean bias direction similarly for both latent classes. The conclusion applied to different MD levels, presence of correlated ME, and combinations of item factor loadings and intercepts. However, the impact on other aspects of growth rates recovery from manipulated L-MI configurations was not always the same between the latent classes. More specifically, contamination %, combination of item intercepts and factor loadings exhibited differential effects on slope factor mean and variance bias/RMSE magnitude between the MD levels. Since no universal conclusion can be made, the discussion on the different between-class impact on the growth rates recovery focuses on the results from F2-GMM as it was the most robust among the models under majority of L-MI deviations.

Under F2-GMM, the impact from inappropriately accounted for L-MI deviation on the slope factor recovery was different for the respective latent class. Slope factors were estimated with acceptable accuracy and high precision in the dominant class but with large bias and low precision in the rare class. As L-MI was evaluated only in one latent class cases in previous studies, this finding in the different impact between the latent classes from the same L-MI deviation was new.

In particular, the rare class was more susceptible to the non-invariant loading of the identification item (i.e., NiLpT condition) than the dominant class. Wirth (2008) found that both slope factor mean and variance estimates had severe bias and low precision when the identification item has non-invariant item parameters in LGM. This conclusion applied to only the rare class in this study with existence of extremely separated latent classes. With non-invariant identification item parameters constrained to be the same in the analysis model, the bias from the incorrectly specified measurement model was expected to be forcefully pushed into the estimates of growth factor scores. Hence, the finding is somewhat counter-intuitive. With the latent classes being extremely well separated, even with the systematically changing item parameters which essentially change the overlap between the latent classes, it was expected to have clear enough interpretation on the recovery of growth factors. With between-class measurement invariance assumed, the same degree of impact from L-MI violation was expected between classes, which is not the case in this study. With a larger sample size in the dominant class, deviation from L-MI assumption was expected to have a more obvious impact as there was more power to detect such violation. Given the above, the reason why the dominant class was not impacted by the non-invariance of the identification item might be related to how the growth factor means (or location of the growth trajectories) were identified during estimation with existence of multiple latent classes. The dominant class might be estimated with more stability given its larger sample size as compared to the rare class. With the latent classes being well separated, the dominant latent class containing a large proportion of individuals did not suffer from deviation of L-MI assumption on its growth rate estimate recovery even when identification invariance failed. However, as there was no literature supporting this finding, more analytical work is needed to find out the underlying reason for this finding.

5.4 LIMITATIONS AND FUTURE RESEARCH

As any other simulation studies, this study inevitably is associated with a few limitations. Since the primary purpose of the study is focused on L-MI violations, majority of the simulated factors were about measurement model characteristics. As indicated from the results, there was the competing effect between latent class distance and changing measurement characteristics. This study is limited in studying the impact of measurement characteristics with fixed mixture proportions and only two latent class distances. While extremely well separated latent classes was used to evaluate the effect of L-MI deviation, another level of MD could have been simulated to be compared to MD = 1.5. Classification accuracy in the study was found to be poor and unrelated to variation of L-MI violation. More balanced proportions could have been simulated to evaluate the recovery of marginal classification (i.e., mixture proportion) then the cell classification (e.g., individual truly generated from the rare class was classified into the rare class). Moreover, as this study is the first study that evaluated L-MI violation on growth characteristics in multiple latent classes, only linear shaped class-specific growth trajectory was included, and the number of latent classes was set as 2 for simplicity in interpretation. In real research practice, there was often class-specific non-linear growth and the number of classes was more often than 2.

The future research could focus more on the impact from the latent class configurations by varying the mixture proportion and the number of latent classes, and adding more levels to the class distance factor. Moreover, non-linear growth pattern can be considered, which increases the complexity in estimating separating classes.

5.5 IMPLICATIONS AND RECOMMENDATIONS

When latent classes were extremely well separated, there was a more distinct pattern on the impact of the measurement characteristics on the growth recovery. F2-GMM is recommended as it was the most robust model with presence of L-MI deviation. Meanwhile, the use of the identification item was critical. There was a need to find the item that was truly invariant in terms of item factor loading and item intercept to avoid the drift of the scale over time. Growth rates can be well recovered when F2-GMM was used. However, researchers should be cautious relying solely on the robustness of the F2-GMM. Unbiasedly recovered growth parameters do not indicate that there is no necessity in a closer and critical scrutiny to the instrument that is used to score the factor. Explicit test on the L-MI assumption is needed. Violation of the L-MI could reflect systematic manifestation of the construct and/or the item parameters changing relationship to the construct over time. It calls for more than a robust model to ensure the reliable use of the instrument, and the valid interpretation on the score from the instrument. Classification was not satisfactory among individuals even under minimal or no L-MI deviation. Researchers should be cautioned in interpreting individual classification using GMMs.

When latent classes were not as well separated which was more common in practice, the growth recovery was challenged as the latent class distance was tangled with non-invariance configuration. The growth rate estimates were marginally recovered by F2-GMM but not the other two models. So, F2-GMM is still recommended. One thing to note is that even though the growth rates were marginally recovered, the Type I error and power by F2-GMM were unacceptable. Hence, the interpretation on the classification is questionable as the clinical meaning in how the growth unfolds over time was altered.

Given the above, appropriate identification of measurement model is critical when latent classes were unobserved. Besides tests to use to find identification item, researchers should strive to find more items that meet L-MI assumption to the maximal degree to minimize the negative impact on the underlying factor scores estimation (i.e., growth parameters).

APPENDIX A

Table A1. Mean raw bias on slope factor mean for C1 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	-0.02	-0.04	-0.01	-0.05	-0.04	-0.01
0%	strong	None	0.02	-0.05	-0.02	-0.02	-0.04	-0.02
50%	pLiT	None	0.02	-0.03	-0.01	-0.01	-0.03	-0.01
50%	iLpT	+	0.03	0.04	-0.02	0.07	0.04	-0.02
50%	pLpT	+	0.11	0.05	-0.01	0.06	0.05	-0.01
100%	pLNiT	+	0.13	0.08	-0.01	0.11	0.09	-0.01
100%	NiLpT	+	0.17	0.04	-0.02	0.14	0.05	-0.01
50%	iLpT	-	-0.07	-0.12	-0.02	-0.12	-0.11	-0.01
50%	pLpT	-	-0.03	-0.11	-0.01	-0.06	-0.11	-0.02
100%	pLNiT	-	-0.11	-0.16	-0.01	-0.07	-0.15	-0.02
100%	NiLpT	-	0.01	-0.1	-0.01	-0.02	-0.11	-0.01

Table A2. Mean raw bias on slope factor mean for C2 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	0.09	0.00	0.00	0.07	0.00	0.00
0%	strong	None	0.10	0.00	0.00	0.08	0.00	0.00
50%	pLiT	None	0.08	-0.01	-0.01	0.06	-0.01	0.00
50%	iLpT	+	0.17	0.05	0.00	0.16	0.04	0.00
50%	pLpT	+	0.15	0.04	0.00	0.13	0.03	0.00
100%	pLNiT	+	0.19	0.06	0.00	0.17	0.05	0.00
100%	NiLpT	+	0.13	-0.07	-0.12	0.10	-0.07	-0.12
50%	iLpT	-	0.03	-0.06	0.00	0.03	-0.06	0.00
50%	pLpT	-	0.01	-0.07	0.00	-0.02	-0.07	0.00
100%	pLNiT	-	-0.04	-0.13	0.00	-0.05	-0.12	0.00
100%	NiLpT	-	-0.02	-0.18	-0.11	-0.04	-0.18	-0.12

Table A3. Mean raw bias on slope factor mean for C1 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	-0.13	-0.08	-0.02	-0.17	-0.08	-0.02
0%	strong	None	-0.06	-0.09	-0.03	-0.12	-0.08	-0.03
50%	pLiT	None	-0.09	-0.06	-0.03	-0.12	-0.06	-0.03
50%	iLpT	+	-0.12	0.01	-0.03	-0.01	0.01	-0.03
50%	pLpT	+	0.02	0.02	-0.02	-0.07	0.02	-0.02
100%	pLNiT	+	0.01	0.05	-0.03	-0.01	0.06	-0.03
100%	NiLpT	+	0.1	0.01	-0.03	0.06	0.02	-0.02
50%	iLpT	-	-0.16	-0.16	-0.04	-0.26	-0.14	-0.03
50%	pLpT	-	-0.13	-0.14	-0.03	-0.17	-0.15	-0.04
100%	pLNiT	-	-0.22	-0.18	-0.02	-0.15	-0.18	-0.04
100%	NiLpT	-	-0.07	-0.13	-0.02	-0.13	-0.14	-0.03

Table A4. Mean raw bias on slope factor mean for C2 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	0.30	-0.04	0.00	0.27	-0.01	0.00
0%	strong	None	0.32	-0.04	0.00	0.31	-0.02	0.01
50%	pLiT	None	0.29	-0.02	0.00	0.31	-0.03	0.01
50%	iLpT	+	0.36	0.04	0.00	0.38	0.05	0.00
50%	pLpT	+	0.39	0.06	0.01	0.37	0.06	0.03
100%	pLNiT	+	0.44	0.08	-0.01	0.42	0.08	0.01
100%	NiLpT	+	0.48	0.02	-0.01	0.41	0.03	-0.02
50%	iLpT	-	0.22	-0.12	-0.02	0.13	-0.13	-0.01
50%	pLpT	-	0.23	-0.12	0.00	0.18	-0.12	-0.02
100%	pLNiT	-	0.23	-0.15	0.00	0.23	-0.15	0.00
100%	NiLpT	-	0.29	-0.15	-0.04	0.28	-0.12	-0.04

Table A5. Mean relative bias on slope factor variance for C1 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	0.89	-0.03	-0.01	1.80	0.00	0.00
0%	strong	None	1.51	0.02	0.00	1.02	-0.03	-0.04
50%	pLiT	None	0.89	-0.09	-0.05	1.40	-0.05	-0.02
50%	iLpT	+	2.00	-0.07	0.02	1.17	-0.06	0.00
50%	pLpT	+	0.82	-0.14	-0.02	1.51	-0.14	-0.04
100%	pLNiT	+	1.27	-0.17	0.00	1.30	-0.13	-0.03
100%	NiLpT	+	1.08	-0.28	-0.26	1.00	-0.33	-0.29
50%	iLpT	-	1.17	-0.04	-0.04	1.67	-0.03	-0.03
50%	pLpT	-	1.35	-0.09	-0.04	1.07	-0.09	-0.02
100%	pLNiT	-	1.59	-0.05	0.00	0.75	-0.11	-0.02
100%	NiLpT	-	0.89	-0.27	-0.26	1.09	-0.24	-0.27

Table A6. Mean relative bias on slope factor variance for C2 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	0.31	-0.01	0.00	0.23	-0.01	-0.01
0%	strong	None	0.35	0.00	0.00	0.31	0.01	0.01
50%	pLiT	None	0.46	-0.03	0.03	0.31	-0.06	0.00
50%	iLpT	+	0.43	-0.05	0.01	0.36	-0.07	0.01
50%	pLpT	+	0.38	-0.14	-0.03	0.34	-0.12	0.01
100%	pLNiT	+	0.43	-0.13	0.01	0.32	-0.14	0.03
100%	NiLpT	+	0.65	-0.31	-0.25	0.37	-0.33	-0.27
50%	iLpT	-	0.40	0.01	0.01	0.43	0.05	0.04
50%	pLpT	-	0.43	-0.06	0.00	0.32	-0.02	0.03
100%	pLNiT	-	0.45	-0.02	0.03	0.29	-0.06	-0.02
100%	NiLpT	-	0.53	-0.30	-0.29	0.39	-0.26	-0.26

Table A7. Mean relative bias on slope factor variance for C1 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	1.58	-0.07	-0.04	3.45	-0.01	0.00
0%	strong	None	2.75	0.06	0.01	1.84	-0.03	-0.07
50%	pLiT	None	1.49	-0.12	-0.08	2.58	-0.04	-0.03
50%	iLpT	+	3.71	-0.08	0.04	2.11	-0.04	0.01
50%	pLpT	+	1.31	-0.20	-0.07	2.79	-0.14	-0.06
100%	pLNiT	+	2.22	-0.19	0.02	2.39	-0.11	-0.06
100%	NiLpT	+	1.72	-0.21	-0.24	1.68	-0.33	-0.31
50%	iLpT	-	2.07	-0.06	-0.06	3.12	-0.06	-0.06
50%	pLpT	-	2.36	-0.12	-0.08	1.91	-0.12	-0.02
100%	pLNiT	-	2.87	-0.06	-0.01	1.25	-0.19	-0.05
100%	NiLpT	-	1.33	-0.28	-0.26	1.86	-0.21	-0.28

Table A8. Mean relative bias on slope factor variance for C2 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	4.22	0.10	0.01	4.00	0.02	0.07
0%	strong	None	4.33	0.05	-0.01	3.64	0.14	0.08
50%	pLiT	None	3.57	0.11	0.08	3.69	0.01	0.02
50%	iLpT	+	4.55	0.08	0.08	3.78	-0.07	-0.02
50%	pLpT	+	4.36	0.07	0.06	2.91	-0.07	0.02
100%	pLNiT	+	4.89	-0.19	-0.06	4.53	-0.08	0.05
100%	NiLpT	+	4.40	-0.18	-0.17	3.57	-0.25	-0.23
50%	iLpT	-	4.20	0.00	0.06	2.70	-0.07	0.00
50%	pLpT	-	3.16	0.04	0.08	2.23	-0.05	-0.03
100%	pLNiT	-	5.46	-0.16	-0.07	3.44	-0.11	-0.05
100%	NiLpT	-	4.48	-0.18	-0.06	4.61	-0.27	-0.19

Table A9. Mean overall classification for C1 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	804	804	804	802	802	802
0%	strong	None	803	803	803	802	802	802
50%	pLiT	None	802	802	802	803	803	802
50%	iLpT	+	803	803	803	802	802	802
50%	pLpT	+	805	805	805	803	802	803
100%	pLNiT	+	802	802	802	803	803	803
100%	NiLpT	+	803	802	802	803	803	803
50%	iLpT	-	802	802	802	802	802	802
50%	pLpT	-	802	802	802	802	802	802
100%	pLNiT	-	803	803	803	804	804	804
100%	NiLpT	-	804	804	804	803	803	803

Table A10. Mean overall classification for C2 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	196	196	196	198	198	198
0%	strong	None	197	197	197	198	198	198
50%	pLiT	None	198	198	198	197	197	198
50%	iLpT	+	197	197	197	198	198	198
50%	pLpT	+	195	195	195	197	198	197
100%	pLNiT	+	198	198	198	197	197	197
100%	NiLpT	+	198	198	198	197	197	197
50%	iLpT	-	198	198	198	198	198	198
50%	pLpT	-	198	198	198	198	198	198
100%	pLNiT	-	197	197	197	196	196	196
100%	NiLpT	-	196	196	196	197	197	197

Table A11. Mean overall classification for C1 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	548	469	653	541	474	704
0%	strong	None	584	451	666	562	462	658
50%	pLiT	None	496	531	662	523	513	650
50%	iLpT	+	495	515	655	561	527	668
50%	pLpT	+	557	565	689	519	557	696
100%	pLNiT	+	523	496	653	510	492	671
100%	NiLpT	+	627	468	641	573	524	687
50%	iLpT	-	544	455	613	452	460	659
50%	pLpT	-	528	488	698	468	483	618
100%	pLNiT	-	539	460	640	585	498	659
100%	NiLpT	-	558	506	687	513	463	658

Table A12. Mean overall classification for C2 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	452	531	347	459	526	296
0%	strong	None	416	549	334	438	538	342
50%	pLiT	None	504	469	338	477	487	350
50%	iLpT	+	505	485	345	439	473	332
50%	pLpT	+	443	435	311	481	443	304
100%	pLNiT	+	477	504	348	490	508	329
100%	NiLpT	+	373	532	359	427	476	313
50%	iLpT	-	456	545	387	548	540	341
50%	pLpT	-	472	513	302	532	517	382
100%	pLNiT	-	461	540	360	415	502	341
100%	NiLpT	-	442	494	313	487	537	342

Table A13. Mean individual classification for C1 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	448	526	417	475	545	385
0%	strong	None	537	619	475	537	595	467
50%	pLiT	None	479	552	494	440	642	440
50%	iLpT	+	545	557	479	514	584	436
50%	pLpT	+	445	604	480	452	568	428
100%	pLNiT	+	498	591	502	413	580	471
100%	NiLpT	+	439	575	448	394	638	494
50%	iLpT	-	564	626	463	513	595	455
50%	pLpT	-	474	602	482	502	614	424
100%	pLNiT	-	521	584	452	480	561	522
100%	NiLpT	-	391	612	464	386	592	479

Table A14. Mean individual classification for C2 under MD = 5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	109	126	102	116	131	95
0%	strong	None	129	148	115	130	143	114
50%	pLiT	None	116	133	120	107	153	108
50%	iLpT	+	131	133	116	124	140	107
50%	pLpT	+	107	143	115	110	136	105
100%	pLNiT	+	121	142	122	101	139	114
100%	NiLpT	+	107	138	110	97	152	119
50%	iLpT	-	136	150	113	124	142	111
50%	pLpT	-	116	144	117	122	147	104
100%	pLNiT	-	125	139	110	116	134	125
100%	NiLpT	-	95	145	112	95	141	116

Table A15. Mean individual classification for C1 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	341	280	341	317	290	320
0%	strong	None	322	267	320	317	273	375
50%	pLiT	None	270	317	330	317	279	311
50%	iLpT	+	259	326	375	311	327	318
50%	pLpT	+	309	304	372	283	307	378
100%	pLNiT	+	279	282	370	290	314	341
100%	NiLpT	+	377	259	326	327	312	354
50%	iLpT	-	303	289	298	245	298	312
50%	pLpT	-	302	282	330	264	291	319
100%	pLNiT	-	337	276	343	331	285	350
100%	NiLpT	-	329	294	360	293	264	367

Table A16. Mean individual classification for C2 under MD = 1.5

			No Correlated ME			Correlated ME		
Contamination	L-MI Pattern	Directional Change on Item Intercepts	1-GMM	C2-GMM	F2-GMM	1-GMM	C2-GMM	F2-GMM
0%	strict	None	215	169	182	246	126	161
0%	strong	None	220	157	182	231	146	175
50%	pLiT	None	295	129	177	246	133	177
50%	iLpT	+	294	128	169	253	135	171
50%	pLpT	+	242	130	147	252	115	149
100%	pLNiT	+	253	142	179	245	151	179
100%	NiLpT	+	199	146	174	218	122	161
50%	iLpT	-	241	141	206	279	139	183
50%	pLpT	-	256	141	148	253	145	189
100%	pLNiT	-	238	153	178	213	134	179
100%	NiLpT	-	259	141	156	247	154	212

BIBLIOGRAPHY

Allan, N. P., Capron, D. W., Lejuez, C. W., Reynolds, E. K., MacPherson, L., & Schmidt, N. B. (2014). Developmental Trajectories of Anxiety Symptoms in Early Adolescence: The Influence of Anxiety Sensitivity. *Journal of abnormal child psychology, 42*(4), 589-600. doi: 10.1007/s10802-013-9806-0

Bauer, D. J., & Curran, P. J. (2003). Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychological methods, 8*(3), 338-363. doi: 10.1037/1082-989x.8.3.338

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

Bongers, I. L., Koot, H. M., Jan van der, E., & Verhulst, F. C. (2004). Developmental Trajectories of Externalizing Behaviors in Childhood and Adolescence. *Child development, 75*(5), 1523-1537. doi: 10.1111/j.1467-8624.2004.00755.x

Bowers, A., & Sprott, R. (2012). Examining the Multiple Trajectories Associated with Dropping Out of High School: A Growth Mixture Model Analysis. *The Journal of Educational Research, 105*(3), 176-195. doi: 10.1080/00220671.2011.552075

Broadbent, J. M., Dickson, N., Moffitt, T. E., Harrington, H., Thomson, W. M., Hancox, R. J., . . . Poulton, R. (2007). Prediction of Differential Adult Health Burden by Conduct Problem Subtypes in Males. *Archives of General Psychiatry, 64*(4), 476-484. doi: 10.1001/archpsyc.64.4.476

Byrne, B. M. (1988). The Self Description Questionnaire III: Testing for Equivalent Factorial Validity across Ability. *Educational and Psychological Measurement, 48*(2), 397-406. doi: 10.1177/0013164488482012

Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian non-clinical adolescents. *Personality and Individual Differences, 21*(5), 641-651. doi: 10.1016/0191-8869(96)00134-1

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological bulletin, 105*(3), 456-466. doi: 10.1037/0033-2909.105.3.456

Callina, K. S., Johnson, S. K., Buckingham, M. H., & Lerner, R. M. (2014). Hope in Context: Developmental Profiles of Trust, Hopeful Future Expectations, and Civic Engagement Across Adolescence. *Journal of Youth and Adolescence, 43*(6), 869-883. doi: 10.1007/s10964-014-0096-9

Caspi, A., & Roberts, B. W. (2001). Personality Development Across the Life Course:The Argument for Change and Continuity. *Psychological Inquiry, 12*(2), 49-66. doi: 10.1207/s15327965pli1202_01

Chan, D. (1997). Racial Subgroup Differences in Predictive Validity Perceptions on Personality and Cognitive Ability Tests. *Journal of Applied Psychology, 82*(2), 311-320. doi: 10.1037/0021-9010.82.2.311

Chan, D. (1998). The Conceptualization and Analysis of Change Over Time: An Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS) and Multiple Indicator Latent Growth Modeling (MLGM). *Organizational Research Methods, 1*(4), 421-483. doi: 10.1177/109442819814004

Chen, 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, Q., Kwok, O.-M., Willson, V., & Luo, W. (2010). The Impact of Ignoring a Level of Nesting Structure in Multilevel Growth Mixture Models: A Monte Carlo Study. *Structural Equation Modeling: A Multidisciplinary Journal, 17*(4), 570-589. doi: 10.1080/10705511.2010.510046

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's Corner: Testing Measurement Invariance of Second-Order Factor Models. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(3), 471-492. doi: 10.1207/s15328007sem1203_7

Collins, J. M., & Gleaves, D. H. (1998). Race, Job Applicants, and the Five-Factor Model of Personality: Implications for Black Psychology, Industrial/Organizational Psychology, and the Five-Factor Theory. *Journal of Applied Psychology, 83*(4), 531-544. doi: 10.1037/0021-9010.83.4.531

Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70*(4), 976-985. doi:10.1037//0022-006X.70.4.976

Dekker, M. C., Ferdinand, R. F., Van Lang, N. D. J., Bongers, I. L., Van Der Ende, J., & Verhulst, F. C. (2007). Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. *Journal of Child Psychology and Psychiatry, 48*(7), 657-666. doi: 10.1111/j.1469-7610.2007.01742.x

Depaoli, S. (2010). *Specification issues in Bayesian growth mixture modeling.* Dissertation/Thesis, ProQuest, UMI Dissertations Publishing. Retrieved from http://pitt.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV3JTsMwEB0hThVIQNkXyT8QaGLHsU9ILBUf0HvlydhQ0ZatSPTvGTsJQq24cLR8skd6s78HIIvLQbaCCR6dJna9odJ16XVN6LBGXZDTFbtYXBmy7FghWmt3IJmQm17qWDS_MnpgC3ZPxfXrWxZVpGK3tZXUYEhmbFZN73Y1ee_OeVwqLa1N-kGFKjlRlLplgerOcg2ik98Z7kA3QpromSY0-9meXqdz_O8rdmHr7ldTfg82_Lwf1ZzbyY8-9GJE2hA674NJivWhLfWJZLcPMZmLG7f0cSNTPHJmv3gSs8lX7E6IpLXDDvIARsP70e1D1sovZM8cI2X1gHKp0HlO-IwjE6oqD7pEa4ijJIq8-tLF8EhqqXPvLAYbrHXkooRJcPIQtl2c0p8v0jYfHYMgw_Y2qvBorHJWISFWeU0BUQeGjhM46n5pTNPpmINPxSDBSd3pnzdn0Gs6-rEscg6bi_dPf9Ew1X4D-369rg

www.summon.com

Depaoli, S. (2012). The Ability for Posterior Predictive Checking to Identify Model Misspecification in Bayesian Growth Mixture Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(4), 534-560. doi: 10.1080/10705511.2012.713251

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus bayesian estimation. *Psychological methods, 18*(2), 186-219. doi: 10.1037/a0031609

Depaoli, S., & Boyajian, J. (2014). Linear and nonlinear growth models: describing a Bayesian perspective. *Journal of consulting and clinical psychology, 82*(5), 784-802. doi: 10.1037/a0035147

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological bulletin, 95*(1), 134-135. doi: 10.1037/0033-2909.95.1.134

Drasgow, F. (1987). Study of the Measurement Bias of Two Standardized Psychological Tests. *Journal of Applied Psychology, 72*(1), 19.

DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science, 46*(1), 137.

Eid, M., & Diener, E. (2006). *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association,

Enders, C., & Tofighi, D. (2008). The Impact of Misspecifying Class-Specific Residual Variances in Growth Mixture Models. *Structural Equation Modeling: A Multidisciplinary Journal, 15*(1), 75. doi: 10.1080/10705510701758281

Estabrook, R., Ram, N., & Grimm, K. (2010). Nonlinear Structured Growth Mixture Models in Mplus and OpenMx. *Multivariate Behavioral Research, 45*(6), 887-909. doi: 10.1080/00273171.2010.531230

Fergusson, D. M., Horwood, L. J., Caspi, A., Moffitt, T. E., & Silva, P. A. (1996). The (artifactual) remission of reading disability: Psychometric lessons in the study of stability and change in behavioral development. . *Developmental Psychology, 32*(1), 132-140. doi: 10.1037/0012-1649.32.1.132

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial Invariance and the Specification of Second-Order Latent Growth Models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 4*(1), 22-36. doi: 10.1027/1614-2241.4.1.22

Galatzer-Levy, I. (2011). *Mapping longitudinal trajectories of adjustment throughout college: A latent growth mixture model approach.* (AAI3448374). Retrieved from http://pitt.idm.oclc.org/login?url=http://search.proquest.com/docview/916525285?accountid=14709

http://RT4RF9QN2Y.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/PsycINFO&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&rft.genre=dissertations+%26+theses&rft.jtitle=&rft.atitle=&rft.au=Galatzer-Levy%2C+Isaac&rft.aulast=Galatzer-Levy&rft.aufirst=Isaac&rft.date=2011-01-01&rft.volume=&rft.issue=&rft.spage=&rft.isbn=9781124540085&rft.btitle=&rft.title=Mapping+longitudinal+trajectories+of+adjustment+throughout+college%3A+A+latent+growth+mixture+model+approach.&rft.issn=&rft_id=info:doi/ PsycINFO database.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs. *The Journal of applied behavioral science, 12*(2), 133-157. doi: 10.1177/002188637601200201

Gunn, J., Elliott, P., Densley, K., Middleton, A., Ambresin, G., Dowrick, C., . . . Griffiths, F. (2013). A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. *Journal of affective disorders, 148*(2-3), 338.

Hancock, G. R., Stapleton, L. M., & Arnold-Berkovits, I. (2005). The tenuousness of invariance tests within multisample covariance structure models. *Unpublished manuscript.*

Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.

Hertzog, C., & Schaie, K. W. (1986). Stability and Change in Adult Intelligence: 1. Analysis of Longitudinal Covariance Structures. *Psychology and aging, 1*(2), 159-171. doi: 10.1037/0882-7974.1.2.159

Hix-Small, H., Duncan, T. E., Duncan, S. C., & Okut, H. (2004). A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology and Behavioral Assessment, 26*(4), 255-270. doi:10.1023/B:JOBA.0000045341.56296.fa

Hong, S. B., Youssef, G. J., Song, S. H., Choi, N. H., Ryu, J., McDermott, B., . . . Kim, B. N. (2014). Different clinical courses of children exposed to a single incident of psychological trauma: a 30-month prospective follow-up study. *Journal of Child Psychology and Psychiatry, 55*(11), 1226-1233. doi: 10.1111/jcpp.12241

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3-4), 117-144.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409-426. doi: 10.1007/BF02291366

Kim, S.-Y. (2012). Sample Size Requirements in Single- and Multiphase Growth Mixture Models: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(3), 457-476. doi: 10.1080/10705511.2012.687672

Kirves, K., Kinnunen, U., De Cuyper, N., & Mäkikangas, A. (2014). Trajectories of perceived employability and their associations with well-being at work: A three-wave study. *Journal of Personnel Psychology, 13*(1), 46-57. doi: http://dx.doi.org/10.1027/1866-5888/a000103

Kohli, N., Harring, J. R., & Hancock, G. R. (2013). Piecewise Linear–Linear Latent Growth Mixture Models With Unknown Knots. *Educational and Psychological Measurement, 73*(6), 935-955. doi: 10.1177/0013164413496812

Landa, R. J., Gross, A. L., Stuart, E. A., & Bauman, M. (2012). Latent class analysis of early developmental trajectory in baby siblings of children with autism. *Journal of Child Psychology and Psychiatry, 53*(9), 986-996. doi: 10.1111/j.1469-7610.2012.02558.x

Lane, S. P., Bluestone, C., & Burke, C. T. (2013). Trajectories of BMI from early childhood through early adolescence: SES and psychosocial predictors. *British Journal of Health Psychology, 18*(1), 66-82. doi: 10.1111/j.2044-8287.2012.02078.x

Lavender, J. M., Kyle, P. D. Y., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Mitchell, J. E., . . . Grange, D. L. (2013). Daily patterns of anxiety in anorexia nervosa: Associations with eating disorder behaviors in the natural environment. *Journal of abnormal psychology, 122*(3), 672.

Leite, W. (2007). A Comparison of Latent Growth Models for Constructs Measured by Multiple Items. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 581-610. doi: 10.1080/10705510701575438

Liu, J. (2012). *A systematic investigation of within-subject and between-subject covariance structures in growth mixture models.* (AAI3553309). Retrieved from http://pitt.idm.oclc.org/login?url=http://search.proquest.com/docview/1492668496?accountid=14709

http://RT4RF9QN2Y.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/PsycINFO&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&rft.genre=dissertations+%26+theses&rft.jtitle=&rft.atitle=&rft.au=Liu%2C+Junhui&rft.aulast=Liu&rft.aufirst=Junhui&rft.date=2013-01-01&rft.volume=&rft.issue=&rft.spage=&rft.isbn=9781267927132&rft.btitle=&rft.title=A+systematic+investigation+of+within-subject+and+between-

subject+covariance+structures+in+growth+mixture+models.&rft.issn=&rft_id=info:doi/ PsycINFO database.

Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement, 74*(4), 557-584. doi:10.1177/0013164413519798

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass U6 - ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info:sid/summon.serialssolutions.com&rft_val_fmt=info:ofi/fmt:kev:mtx:book &rft.genre=book&rft.title=Statistical+theories+of+mental+test+scores&rft.au=Lord%2C +Frederic+M&rft.au=Novick%2C+Melvin+R&rft.series=The+Addison-Wesley+series+in+behavioral+science%3A+quantitative+methods&rft.date=1968&rft.pub=Addison-Wesley+Pub.+Co&rft.externalDocID=12891&paramdict=en-US U7 - Book: Addison-Wesley Pub. Co.

Lu, Z. L. (2011). *Bayesian inference of robust growth mixture models with non-ignorable missing data.* Dissertation/Thesis, ProQuest, UMI Dissertations Publishing. Retrieved from http://pitt.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV3JasMwEBWlp9BC23RfQD_gVpZkST4VuoSeS-5BluTEkNht4kDz953xEkpKLwVfBh2MLfFmnmbmDSGC37NoBxOY9eC6Daq3JUq7AI43DRK8j7Xc6tjtFFn2qhDdbvcg2SC3rxxemj-Y1EAorhL--PEZ4RQpzLZ2IzUAkgGbZZu73SXvvY2xjlSGNfODgIVAaKE7TZ6trX9BdON3RkekLyFt5JkKv9h2T_-Wc_zvVxyTg5cfSfkTshfKIU5z7io_hmSwRckNGBieturOp-T9yW4CtmHSom8cpFVOl1W2XtV0ChS_ntFF8YVpCtoM3VlRvPmlZVVGxbSE85fNYQXeDy6UYrnqGRmPXsfPb1E3pSGaySSOFO6lDEYJkamcW5Vri5pyQmsmdc68l9YkyqU8s7mVUqbeMStSGeAsaCfFOTm0WMxf1k3Tn78kVHrgWbH3ikugr8Ea441LQ2atEC5nyRW56H_mxM_nEw6siwN3lfH1nys3ZNBeB-NzS_br5TrctYK2327CyV8

www.summon.com

Lubke, G. H., & Dolan, C. V. (2003). Can Unequal Residual Variances Across Groups Mask Differences in Residual Means in the Common Factor Model? *Structural Equation Modeling: A Multidisciplinary Journal, 10*(2), 175-192. doi: 10.1207/S15328007SEM1002_1

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., . . . Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of consulting and clinical psychology, 82*(2), 287-297. doi: 10.1037/a0035535

Lydecker, K. P. (2011). *Role of gender in the presentation and course of antisocial behavior for alcohol and drug-disordered youth.* (AAI3426046). Retrieved from http://pitt.idm.oclc.org/login?url=http://search.proquest.com/docview/873851683?accountid=14709 http://RT4RF9QN2Y.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/PsycINFO&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&rft.genre=dissertations+%26+theses&rft.jtitle=&rft.atitle=&rft.au=Lydecker%2C+Katherine+Patterson&rft.aulast=Lydecker&rft.aufirst=Katherine&rft.date=2011-01-01&rft.volume=&rft.issue=&rft.spage=&rft.isbn=9781124260563&rft.btitle=&rft.title=Role+of+gender+in+the+presentation+and+course+of+antisocial+behavior+for+alcohol+and+drug-disordered+youth.&rft.issn=&rft_id=info:doi/ PsycINFO database.

MacCallum, R. C. (2009). Factor Analysis. In R. E. Millsap & A. Maydeu-Olivares (Eds.), The SAGE Handbook of Quantitative Methods in Psychology (pp.123- 147), Thousand Oaks, CA: Sage Publications.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed., pp. 561–614). New York: Plenum.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Meade, A. W., & Lautenschlager, G. J. (2004). A Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance. *Organizational Research Methods, 7*(4), 361-388. doi: 10.1177/1094428104268027

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543. doi: 10.1007/BF02294825

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122. doi: 10.1007/BF02294746

Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research, 30*(4), 577-605. doi: http://dx.doi.org/10.1207/s15327906mbr3004_6

Millsap, R. E. (1997). Invariance in Measurement and Prediction: Their Relationship in the Single-Factor Case. *Psychological methods, 2*(3), 248-260. doi: 10.1037/1082-989X.2.3.248

Millsap, R. E., & Hartog, S. B. (1988). Alpha, Beta, and Gamma Change in Evaluation Research: A Structural Equation Approach. *Journal of Applied Psychology, 73*(3), 574-584. doi: 10.1037/0021-9010.73.3.574

Moffitt, T. E. (1993). Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy. *Psychological review, 100*(4), 674-701. doi: 10.1037/0033-295x.100.4.674

Mustillo, S. A., Hendrix, K. L., & Schafer, M. H. (2012). Trajectories of Body Mass and Self-Concept in Black and White Girls: The Lingering Effects of Stigma. *Journal of health and social behavior, 53*(1), 2-16. doi: 10.1177/0022146511419205

Muthén, B., & Shedden, K. (1999). Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics, 55*(2), 463-469. doi: 10.1111/j.0006-341X.1999.00463.x

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882-891. doi: 10.1097/00000374-200006000-00020

Muthén, L. K., & Muthén, B. (1998). Mplus. Statistical analyses with latent variables. Los Angeles, CA: Muthén & Muthén.

Muthén, B., Asparouhov, T., & Nylund, K. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569. doi: 10.1080/10705510701575396

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C., . . . Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics (Oxford, England), 3*(4), 459-475. doi: 10.1093/biostatistics/3.4.459

Neale, M. C., & Lubke, G. (2006). Distinguishing Between Latent Classes and Continuous Factors: Resolution by Maximum Likelihood? *Multivariate Behavioral Research, 41*(4), 499-532. doi: 10.1207/s15327906mbr4104_4

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and psychopathology, 20*(2), 673-716. doi: 10.1017/s0954579408000333

Oh, W., Rubin, K. H., Bowker, J. C., Booth-LaForce, C., Rose-Krasnor, L., & Laursen, B. (2008). Trajectories of Social Withdrawal from Middle Childhood to Early Adolescence. *Journal of abnormal child psychology, 36*(4), 553-566. doi: 10.1007/s10802-007-9199-z

Olivera-Aguilar, M. (2013). *Impact of Violations of Longitudinal Measurement Invariance in Latent Growth Models and Autoregressive Quasi-simplex Models.* Dissertation/Thesis, ProQuest, UMI Dissertations Publishing. Retrieved from http://pitt.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV3LSsNAFB3EVVHw_RbmB6JNJpnprFTUaqEuhOK2zMxNMFiTaltR_HnvnSZVWt1IVpNkEZJw5p77OIcxEZ00gzlMCGn41RjVMk4KBxSFQKhdaHVGmmTRXJPleZ0pmH7tGiQ9ckPpKGl-iht9U7cSqeKz4UtANlJUbq08NRCTEZzjafH2F_ZOyB0h-VOqkn2q13IBk_1G015jdS-A12PK4Xk2Lr2o3_jvx15nK1c_yvAbbCktNlljhoUfW-yz46cneZnxh7ys-uVo1S3J4GgCZKbF774TjLxTvCHtpn-I5wXvYgiLJ2-Q5I8fORmuDUbcFMAvSDMh9SQfcZbfT8woD0Y5aRS_V_dts177und5G1Q-DcETBlNBksoIXAgYCfqqp7F4SIzcECu0colSVisrZQsiIIoHVmBYFDYzK_DeVIsdtmq onb8Y-7E_2GM8c5kTAtlUBpYk5WkA2KSxNonUoMJ4n-3Wb7cPg0EfiY_EyCvW4cGfVw5ZI_JeFpQ_OWLL49dJejyVtP0CcdTNQg

www.summon.com

Pajer, K. A. (1998). What happens to 'bad' girls? A review of the adult outcomes of antisocial adolescent girls. *American Journal of Psychiatry, 155*(7), 862-870. doi: 10.1176/ajp.155.7.862

Peters, B. C., & Walker, H. F. (1978). An Iterative Procedure for Obtaining Maximum-Likelihood Estimates of the Parameters for a Mixture of Normal Distributions. *SIAM Journal on Applied Mathematics, 35*(2), 362-378. doi: 10.1137/0135032

Peugh, J., & Fan, X. (2012). How Well Does Growth Mixture Modeling Identify Heterogeneous Growth Trajectories? A Simulation Study Examining GMM's Performance Characteristics. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 204. doi: 10.1080/10705511.2012.659618

Redner, R. A., & Walker, H. F. (1984). Mixture Densities, Maximum Likelihood and the Em Algorithm. *SIAM Review, 26*(2), 195-239. doi: 10.1137/1026034

Ram, N., & Grimm, K. (2009). A Second-Order Growth Mixture Model for Developmental Research. *Research in Human Development, 6*(2), 121-143. doi: 10.1080/15427600902911221

Ramo, D. E., Anderson, K. G., & Tate, S. R. (2005). Characteristics of relapse to substance use in comorbid adolescents. *Addictive Behaviors [H.W. Wilson - SSA], 30*(9), 1811.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological bulletin, 114*(3), 552. doi: 10.1037/0033-2909.114.3.552

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: lexical development from 2;0 to 3;0. *Journal of child language, 27*(2), 293-311. doi: 10.1017/s030500090000413x

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology, 39*(6), 1020.

Schaubroeck, J., & Green, S. G. (1989). Confirmatory Factor Analytic Procedures for Assessing Change During Organizational Entry. *Journal of Applied Psychology, 74*(6), 892-900. doi: 10.1037/0021-9010.74.6.892

Schumm, J. A., Walter, K. H., & Chard, K. M. (2013). Latent class differences explain variability in PTSD symptom changes during cognitive processing therapy for veterans.

Psychological Trauma: Theory, Research, Practice, and Policy, 5(6), 536-544. doi: http://dx.doi.org/10.1037/a0030359

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research, 25*(1), 78-107. doi: 10.1086/209528

Steinmetz, H. (2011). Analyzing Observed Composite Differences Across Groups: Is Partial Measurement Invariance Enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*(1), 1-12. doi: 10.1027/1614-2241/a000049

Thompson, M. P., Swartout, K. M., & Koss, M. P. (2013). Trajectories and predictors of sexually aggressive behaviors during emerging adulthood. *Psychology of Violence, 3*(3), 247-259. doi: http://dx.doi.org/10.1037/a0030624

Thurstone, L. L. (1947). *Thurstone Interest Schedule*: Psychological Corporation, San Antonio, TX.

Tucker, J. S., Orlando, M., & Ellickson, P. L. (2003). Patterns and Correlates of Binge Drinking Trajectories From Early Adolescence to Young Adulthood. *Health Psychology, 22*(1), 79-87. doi: 10.1037/0278-6133.22.1.79

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 3*(1), 4-70. doi: 10.1177/109442810031002

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in psychology U6 - ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info:sid/summon.serialssolutions.com&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.atitle=Facing+off+with+Scylla+and+Charybdis%3A+a+comparison+of+scalar%2C+partial%2C+and+the+novel+possibility+of+approximate+measurement+invariance&rft.jtitle=Frontiers+in+Psychology&rft.au=van+de+Schoot%2C+Rens&rft.au=Kluytmans%2C+Anouck&rft.au=Tummers%2C+Lars&rft.au=Lugtig%2C+Peter&rft.date=2013&rft.eissn=1664-1078&rft.volume=4&rft_id=info:doi/10.3389%2Ffpsyg.2013.00770&rft.externalDBID=n%2Fa&rft.externalDocID=10_3389_fpsyg_2013_00770&paramdict=en-US U7 - Journal Article, 4*, 770. doi: 10.3389/fpsyg.2013.00770

Wang, M., & Bodner, T. E. (2007). Growth Mixture Modeling: Identifying and Predicting Unobserved Subpopulations With Longitudinal Data. *Organizational Research Methods, 10*(4), 635-656. doi: 10.1177/1094428106289397

Warner, L. A., White, H. R., & Johnson, V. (2007). Alcohol initiation experiences and family history of alcoholism as predictors of problem-drinking trajectories. *Journal of studies on alcohol and drugs, 68*(1), 56-65.

White, H. R., Johnson, V., & Buyske, S. (2000). Parental modeling and parenting behavior effects on offspring alcohol and cigarette use: A growth curve analysis. *Journal of substance abuse, 12*(3), 287-310. doi: 10.1016/s0899-3289(00)00056-0

White, H. R., & Labouvie, E. W. (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol, 50*(1), 30-37.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain (pp. 281-324): American Psychological Association.

Wirth, R. J. (2008). *The effects of measurement non-invariance on parameter estimation in latent growth models.* Dissertation/Thesis, ProQuest, UMI Dissertations Publishing. Retrieved from http://pitt.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV3JTsMwELUQpwok9h3JPxCIl9rxiQNQ8QG9cKq8BSpah6Ug9e8ZO3FBqbhwtCwfEltvtjdvEGL0qix6mMCiW2qJ5ooJoXzlhFa0lCrWsYwytEeyzKoQ3W1nkEzI7Robk-bXscTHOADwzetbEadIxWprN1IDIBmwmbe1237wntdEMrDutEodrFxVFVj2rMmT12oNopPdGe2gTCFN8kxTN191T6_LOf73K3bR1t2vovwe2vBhHw1WyLg8QI_wmHDH_MBNjec_iUUcmlBMwxeE3PH94CbgqCY-jywbHBU82tZIPA14Bm4tHHiCwH_xjNMIno9DNB7dj28fim4mQ_HCh6Qw3ilpqKWOMOME0ZYbIbimQteSWEU0OIhVrWm8bW5IZfWQay_rJFSnLTtC2zpS98Mitfi5E4QN-HyaWmt4bbnzUitXl760QpQWQvfhKTrOv27iZrMJeJBMSqY4Oftz5xwNWmJHzJVcoM3F-6e_bOVrvwHSBsVL

www.summon.com

Wu, J., Witkiewitz, K., McMahon, R. J., Dodge, K. A., & Conduct Problems Prevention Research, G. (2010). A parallel process growth mixture model of conduct problems and substance use with risky sexual behavior. *Drug and alcohol dependence, 111*(3), 207-214. doi: 10.1016/j.drugalcdep.2010.04.013

Yaroslavsky, I., Pettit, J. W., Lewinsohn, P. M., Seeley, J. R., & Roberts, R. E. (2013). Heterogeneous trajectories of depressive symptoms: adolescent predictors and adult outcomes. *Journal of affective disorders, 148*(2-3), 391-399. doi: 10.1016/j.jad.2012.06.028

Zerwas, S., Von Holle, A., Watson, H., Gottfredson, N., & Bulik, C. M. (2014). Childhood anxiety trajectories and adolescent disordered eating: Findings from the NICHD Study of Early Child Care and Youth Development. *International Journal of Eating Disorders, 47*(7), 784-792. doi: http://dx.doi.org/10.1002/eat.22318

Zhang, Z., Lu, Z., & Lubke, G. (2011). Bayesian Inference for Growth Mixture Models with Latent Class Dependent Missing Data. *Multivariate Behavioral Research, 46*(4), 567-597. doi: 10.1080/00273171.2011.589261

Zopluoglu, C., Harring, J. R., & Kohli, N. (2014). FitPMM: An R Routine to Fit Finite Mixture of Piecewise Mixed-Effect Models With Unknown Random Knots. *Applied Psychological Measurement, 38*(7), 583-584. doi: 10.1177/0146621614540482