COMPARING MODEL-BASED AND DESIGN-BASED STRUCTURAL EQUATION MODELING APPROACHES IN ANALYZING COMPLEX SURVEY DATA

A Dissertation

by

JIUN-YU WU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Educational Psychology

Comparing Model-based and Design-based Structural Equation Modeling Approaches in

Analyzing Complex Survey Data

COMPARING MODEL-BASED AND DESIGN-BASED STRUCTURAL EQUATION MODELING APPROACHES IN ANALYZING COMPLEX SURVEY DATA

A Dissertation

by

JIUN-YU WU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee, Victor L. Willson
                       Oi-man Kwok
Committee Members,   Bruce Thompson
                       Michael Speed
Head of Department,   Victor L. Willson

August 2010

Major Subject: Educational Psychology

ABSTRACT

Comparing Model-based and Design-based Structural Equation Modeling Approaches in

Analyzing Complex Survey Data. (August 2010)

Jiun-Yu Wu, B.S.; M.S., National Chiao Tung University, Taiwan

Co-Chairs of Advisory Committee: Dr. Victor L. Willson
Dr. Oi-man Kwok

Conventional statistical methods assuming data sampled under simple random sampling are inadequate for use on complex survey data with a multilevel structure and non-independent observations. In structural equation modeling (SEM) framework, a researcher can either use the ad-hoc robust sandwich standard error estimators to correct the standard error estimates (Design-based approach) or perform multilevel analysis to model the multilevel data structure (Model-based approach) to analyze dependent data.

In a cross-sectional setting, the first study aims to examine the differences between the design-based single-level confirmatory factor analysis (CFA) and the model-based multilevel CFA for model fit test statistics/fit indices, and estimates of the fixed and random effects with corresponding statistical inference when analyzing multilevel data.

Several design factors were considered, including: cluster number, cluster size, intra-class correlation, and the structure equality of the between-/within-level models. The performance of a maximum modeling strategy with the saturated higher-level and true lower-level model was also examined. Simulation study showed that the design-based approach provided adequate results only under equal between/within structures. However, in the unequal between/within structure scenarios, the design-based approach produced biased fixed and random effect estimates. Maximum modeling generated consistent and unbiased within-level model parameter estimates across three different scenarios.

Multilevel latent growth curve modeling (MLGCM) is a versatile tool to analyze the repeated measure sampled from a multi-stage sampling. However, researchers often adopt latent growth curve models (LGCM) without considering the multilevel structure. This second study examined the influences of different model specifications on the model fit test statistics/fit indices, between/within-level regression coefficient and random effect estimates and mean structures. Simulation suggested that design-based MLGCM incorporating the higher-level covariates produces consistent parameter estimates and statistical inferences comparable to those from the model-based MLGCM and maintain adequate statistical power even with small cluster number.

DEDICATION

This document is dedicated to Michelle and Katherine, my beautiful and precious

baby girls.

ACKNOWLEDGEMENTS

It is my most honors to be a member of Research, Measurement, and Statistics

Program at Texas A&M University. In this program, I have not only equipped myself with

the professional knowledge for being a quantitative methodologist and psychometrician,

but also cultivated myself to be a cautious and rigorous scientist with human touch. My

dear co-chairs, Dr. Victor Willson and Dr. Oi-man Kwok, have the most impact on my

research and interpersonal skill. These two gentlemen are always kind to share with me

their brilliant ideas, harsh but useful comments on my research, and also personal

experience in academia. Besides that, Dr. Willson also lets me know what the spirit of

cowboy is, that is, never acting like a European soccer player. Like an older brother, Dr.

Oi-man Kwok, who introduces me the beauty of linear modeling and statistical research,

shows me in person how brave, enthusiastic, and sedulous a promising young scholar can

be in academia. I also learn a lot from him the attitude of The Doctrine of the Mean. I am

also thankful for my committee members Dr. Bruce Thompson and Dr. Mike Speed. I was

totally amazed by Dr. Thompson's special teaching method in EPSY 640, no writing but

just speaking, which effectively guided me through the beauty of fundamental statistics.

Even now, I still sometimes hear his voice in my dream. Nevertheless, Dr. Thompson always tries his best to complete my portfolio and point out my weakness. Dr. Speed guides me through the advanced mixed effects modeling and the advanced SAS programming. He has never hesitated to solve my question and point out little the bug(s) from thousands of lines of SAS syntax. I am so lucky to have these four gentlemen in my committee, I admire your insight and passion in quantitative research, and enjoy your brilliant and enlightening comments in my research. You are my most wonderful advisers.

With my sincerest thankfulness, Dr. Jan Hughes, my dear mentor, is one of the most successful scholars with global impact on school psychology, who also gives full respect to her colleagues and, what's more, is surprisingly humble. It is my greatest honor to work for Dr. Hughes in Project Achieve. Under her guidance and supervision, I learn how to work on substantial area research and complete a project as a team. She gives me the flexibility and opportunities to apply what I have learned, and always reminds me the importance of linking methodology to the real world research. Working with her is the most enjoyable experience in my research life. You are my role model in all aspects.

I would like to thank my parents who foster me with their whole heart, attention, forgiveness, and, of course, funding to provide me the best education and even support me

to study abroad. I love you. Special thanks to my parents-in-law. They come to help us

when we have our second baby girl. With their help, I can concentrate on finishing my

dissertation, passing my final defense and job hunting.

Many thanks to my friends in RMS: Hsien-Yuan, Susan, Russell, Ross, Qi, Yan,

June, Eun Sook, Minjung and everyone in RMS program. I will always remember your

help with my questions and the wonderful RMS potlucks we had together. We are the best

quantitative team in the world. Also thanks my dear Taiwanese friends: Peter/Vivian,

Jack/Jenny, Stephan/Rosy, Terence, Shawn, Prudence, Calvin/Erin, YY/CH, John/Jasmine,

Mark/Yenlin and many of you. We come from the same beautiful island, Taiwan, and

strengthen our competitive ability together in this nation, USA. You gave me unselfish

help and strength to get through the ups and downs.

Finally, I am truly grateful for Karen Yuan-Hsuan Lee, my dear wife, and Michelle

and Katherine, my dear baby girls. You are the treasure of my life. With your company, I

never feel lonely and hopeless when I fight against and get through all the difficulties. Yes,

my dear, we really made it for two Ph.D.s and two babies these years in TAMU, College

Station. I cannot tell you how much you complete me. This is just a beginning, and we

will overcome the coming challenges and enjoy our life together. Sincerely.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This journal formatted dissertation consists of one literature review of structural

equation modeling (SEM) techniques commonly used to deal with dependent data

(Section II), and two inter-related simulation studies, one in a cross-sectional setting

(Section III) and the other in a longitudinal setting (Section IV), focusing on issues

associated with the effect of ignoring modeling higher-level variation of dependent data

under the SEM framework. A brief introduction about the issues under concern is

presented below.

The technique of cluster sampling or multistage sampling is widely used in

educational, behavioral and organizational research due to the efficiency in time and

resources. Unlike simple random sampling (SRS) which randomly select a sample from a

target population to assure that the selected observations are independent from each other,

cluster sampling randomly samples naturally occurring groups/clusters of

individuals/observations (Gall, Gall, & Borg, 2006; Stapleton, 2006). Data collected with

_____

This dissertation follows the style of *Psychological Methods.*

the use of cluster sampling are prone to have correlated observations within clusters. For example, students from the same classroom are more likely to respond in a similar way because of the influence from the same environment.

Conventional statistical methods which assume independent observations should not be used with data collected from cluster sampling due to the potential non-independent observations. The use of conventional statistical methods on non-independent data can result in biased estimation of the standard errors and incorrect statistical conclusions (Hox, 2002; Kish, 1995). Multilevel models (Goldstein, 1987,1995), also named hierarchical linear models (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002), random coefficient models (Jennrich & Schluchter, 1986) , random effects models (Laird & Ware, 1982), or covariance component models (Longford, 1993) are preferred strategies to model data of hierarchical structure (Cheung, 2007).

In structural equation modeling (SEM), data are typically assumed to be collected through SRS so that they are independently and identically distributed (Stapleton, 2006; du Toit & du Toit, 2008). In social and educational research, however, it is not ~~rare~~ uncommon to have data with a hierarchical structure, especially when data are obtained through cluster sampling or multistage sampling, where there is dependency among observations such as

students nested within schools or individuals nested within households (Lee & Forthofer,

2006; Skinner, Holt, & Wrigley, 1997). "By ignoring the hierarchical structure of the

data, incorrect parameter estimates, standard errors, and inappropriate fit statistics may be

obtained" (Du Toit & Du Toit, 2008, p.456).

Three analytic approaches are usually used for analyzing data collected through

cluster sampling, namely, disaggregated analysis, aggregated analysis, and multilevel

modeling (Hofmann, 1997). Disaggregated analysis ignores higher level structure of data

(e.g. classroom level) and only models observations at the lower-level structure (e.g.

student level). This approach has been criticized for violating the assumption of

independency e under simple random sampling (Hofmann, 1997; Raudenbush & Bryk,

2002). Neglecting the dependency among observations will generally result in

underestimating the standard errors of the fixed effect and leading to inflated Type I error

rate (De Leeuw & Kreft, 1995; Raudenbush & Bryk, 2002; and Snijders & Bosker, 1999).

On the other hand, aggregated analysis, as its name suggests, only analyzes aggregated

data from the lower/individual level. Studies have shown that regression analysis

performed on aggregated data can result in biased parameter estimates and underestimated

standard errors associated with the fixed effects (Croon & van Veldhoven, 2007; Ludtke et

al., 2008). Moreover, aggregated data cannot fully reflect the individual-level variation as well (Au & Cheung, 2004; Klein, Conn, Smith, & Sorra, 2001). It has been shown that none of these two approaches can adequately reveal the complete picture of the relations between different levels of variables in multilevel data (Holt, Scott, & Ewings, 1980; Raudenbush & Bryk, 2002).

The third approach is to use multilevel modeling, which allows researchers to maintain the original data structure for analyses. There are two common approaches in multilevel modeling, namely, design-based and model-based approaches. The design-based approach takes the multilevel data/dependency into account by adjusting for the standard errors of the parameter estimates based on the sampling design, while the model-based approach analyzes the multilevel data by specifying a level-specific model for each level of the data. For example, for two-level clustered sampling data, the model-based approach analyzes terent) within-level and between-level models respectively whereas the design-based approach analyzes the data with only one overall model and adjusts the standard errors of the parameter estimates based on the sampling design. Nevertheless, the design-based approach is commonly used by substantive researchers (e.g. Agrawal & Lynskey, 2007; Davidov, Yang-Hansen, Gustafsson, Schmidt,

& Bamberg, 2006; Hox & Kleiboer, 2007; Mathews et al., 2009; Muthén & Asparouhov, 2006) given that this approach only requires them to specify one single model, and researchers may mostly be interested in only examining the lower level (or within-level) model.

For analyzing longitudinal data, Latent growth curve modeling (LGCM) is among the multilevel models and draws on many of the strengths under the structural equation modeling (SEM) framework (Curran & Hussong, 2002; Duncan, Duncan, Strycker, Li, & Alpert, 1999). LGCM is capable of analyzing repeated measurement data to provide flexible structural modeling of growth factors, future outcomes of growth, and covariates or constructs to explain the difference in the initial level and trajectories (Duncan & Duncan, 2004; Hancock & Lawrence, 2006; Stoolmiller, 2007). There are several advantages of the LGCM over traditional approaches for analyzing longitudinal data (Curran, 2003; Duncan, Duncan, & Strycker, 2006; Duncan et al., 1999). The emphasis on inter-individual and intra-individual difference, in particular, makes LGCM a popular multivariate statistical method in analyzing longitudinal data (Cheung, 2007). Specifically, researchers have used LGCMs to study changes over time in a longitudinal design; for example, to examine gender differences in the change of academic self-concept and

language achievement (De Fraine, Van Damme, & Onghena, 2007), to investigate

adolescent twin's conflict with their mothers over time (Kashy, Donnellan, Burt, & McGue,

2008) and to study boys' and girls' talent perceptions and intrinsic values through

adolescence (Watt, 2008).

Multilevel latent growth curve model (MLGCM), on the other hand, extends the

concept of LGCM to include cluster-specific higher-level data in the model. Most of the

multilevel data structure originates from the use of multistage sampling or cluster sampling

where the larger sampling units are sampled at the first stage followed by randomly

sampling smaller units within the larger unit (Stapleton, 2006). Under such circumstances,

smaller sampling units within a large sampling unit tend to have similar or dependent

responses due to the influence of the same environment. Ignoring dependency in the

grouped data can cause biased estimates and incorrect statistical inference in the analyses

(D. Kaplan & Elliott, 1997). Dickinson and Basu (2005) suggested that statistical

approaches be used to account for the hierarchical data structure when the data are nested

in nature, or the correct interpretation of the result may be at risk.

However, researchers may fail to run a model that conforms to the original complex

data structure or takes all the data levels into account for the analysis. For instance,

researchers are interested in examining the general reading achievement trajectory over a

large number of students who are from different schools. They may conduct a LGCM to

examine the average growth pattern of the student reading achievement without

considering the school-level effect and assuming that these students are independent of

each other. Possible reasons for this negligence may include cutting down the complexity

in data analysis (Meyers & Beretvas, 2006; Wampold & Serlin, 2000), failing to identify

the primary sampling unit or the higher level identity information (IDs) (Moerbeek, 2004),

and avoiding the nonconvergence issue in model estimation (Van Landeghem, De Fraine,

& Van Damme, 2005).

2. LITERATURE REVIEW OF DESIGN-BASED AND MODEL-BASED

MULTILEVEL TECHNIQUES IN STRUCTURAL EQUATION MODELING

2.1 Structural Equation Modeling

Structural equation modeling (SEM) (Bentler, 1980; Fassinger, 1987; Jöreskog, 1970, 1978) is the most rapidly developing analytic technique over the last three decades and now becomes one of the most commonly used methodologies in various science fields (MacCallum & Austin, 2000). SEM combines two powerful methodologies, path analysis (i.e. structural model which explains the relationship between latent variables) and factor analysis (i.e. measurement model which refines the latent variable from observed variables and is free from measurement error). SEM has been considered as the most general case of the broader parametric General Linear Models (GLM), which takes the measurement error and latent structure into account (e.g. t test, OVA models, multiple regression, and descriptive discriminant analysis, canonical correlation analysis, etc.) (Curran, 2003; Fan, 1997; Graham, 2008; Jöreskog & Sörbom, 1993; Rigdon, 1998; Thompson, 2000). Taking measurement errors into account, SEM avoids the problem of

shrinkage of regression coefficient estimates and provide more accurate estimates of the

structural relationship between observed variables (Heck & Thomas, 2008).

Being the forerunner of SEM, Covariance Structure Analysis (CSA) is the

statistical method of structural analysis with the observed (sample) variance-covariance

matrix (Bock, 1960; Bock & Bargmann, 1966; Jöreskog, 1970; Schmidt, 1969). Jöreskog

(1967, 1969, 1970, 1973, 1977) presented a series of general analytic frameworks of

Covariance Structure Analysis for estimating parameters in latent variable model, also

named linear structural relationship equation system, (i.e. LISREL) (Jöreskog & Sörbom,

1993). SEM is basically the same covariance-based methodology but with the additional

capability to take the mean structure model into account. Based on theories, personal

experiences, and literature reviews, this kind of covariance based methodology allows

researchers to hypothesize their research question in a causality model with multiple

indicators and multiple causes. A set of matrices needs to be specified to represent the

model parameters of structural model. Then the causality model will be summarized in

the model implied mean structure and variance-covariance matrix. By using the fitting

estimation method (e.g. maximum likelihood estimation), the estimates of model

parameter estimates will be calculated by minimizing the discrepancy function between

observed variance-covariance matrix and the model implied one. Finally, several

commonly used model fit likelihood ratio test statistic (i.e. model fit chi-square test

statistic) and fit indices will be provided to evaluate the quality of proposed causality

model.

SEM includes two fundamental blocks of models: measurement model and

structure model as shown in Figure 1.

Measurement Model   Structure Model   Measurement Model

$$\mathbf{X} = \mathbf{u} + \boldsymbol{\Lambda_X}\boldsymbol{\xi} + \boldsymbol{\delta} \qquad \boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad \mathbf{Y} = \mathbf{v} + \boldsymbol{\Lambda_Y}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$
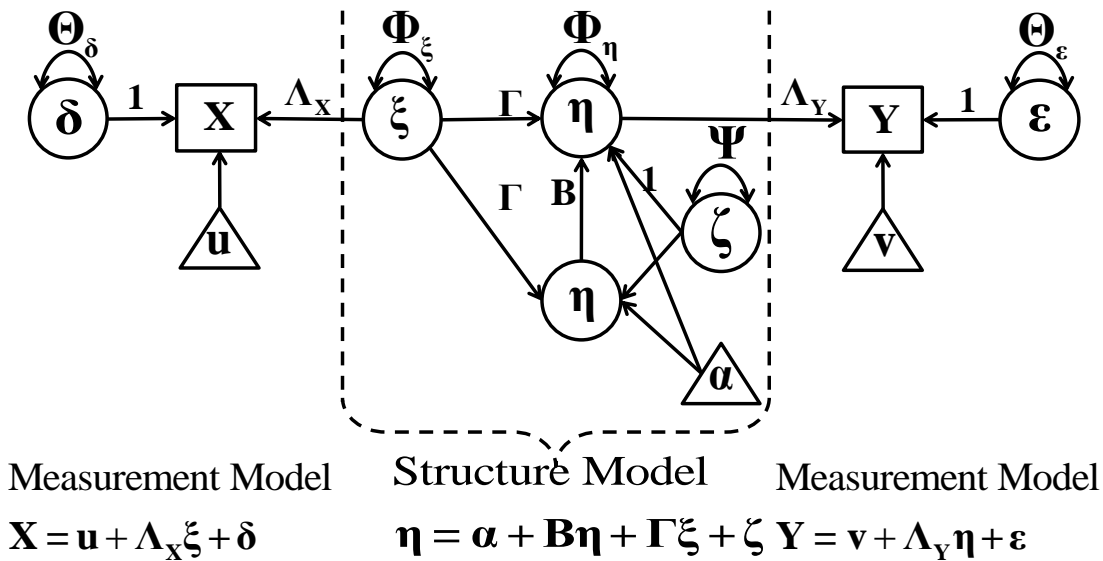
Figure 1. Three latent variables SEM model

The measurement model applies the confirmatory factor analysis (CFA) to adjust

the measurement error of indicators and to form the latent variables (factors). The

measurement model of exogenous indicators **x** and endogenous indicators **y** can be

defined as

$$\mathbf{x} = \mathbf{u} + \mathbf{\Lambda_X}\mathbf{\xi} + \mathbf{\delta} \tag{2.1}$$

$$\mathbf{y} = \mathbf{v} + \mathbf{\Lambda_Y}\mathbf{\eta} + \mathbf{\epsilon} \tag{2.2}$$

The measurement model applies the confirmatory factor analysis (CFA) to adjust the measurement error of indicators and to form the latent variables (factors). The measurement model of exogenous indicators $\mathbf{x}$ and endogenous indicators $\mathbf{y}$ can be defined as

$$\mathbf{x} = \mathbf{u} + \mathbf{\Lambda_X}\mathbf{\xi} + \mathbf{\delta} \tag{2.3}$$

$$\mathbf{y} = \mathbf{v} + \mathbf{\Lambda_Y}\mathbf{\eta} + \mathbf{\epsilon} \tag{2.4}$$

where $\mathbf{x}$ is $Q \times 1$ vector of Q observed exogenous variables x_q, $\mathbf{\xi}$ is the $R \times 1$ vector of R exogenous latent variables ξ_r with $MVN\left(\mathbf{0}, \mathbf{\Phi_\xi}\right)$ (i.e. multivariate normal distribution with mean zero and variance $\mathbf{\Phi_\xi}$), $\mathbf{\Lambda_X}$ is the $Q \times R$ matrix of factor loadings between Q exogenous indicators x_q and R exogenous latent variables ξ_r, $\mathbf{u}$ is $Q \times 1$ vector of Q intercepts u_q, and $\mathbf{\delta}$ is the $Q \times 1$ vector with $MVN\left(\mathbf{0}, \mathbf{\Theta_\delta}\right)$ (i.e. multivariate normal distribution with mean zero and variance $\mathbf{\Theta_\delta}$) of Q observed exogenous unique variables δ_q; $\mathbf{y}$ is $P \times 1$ vector of P observed endogenous variables y_p, $\mathbf{\eta}$ is the $S \times 1$ vector of S endogenous latent variables η_s with $MVN\left(\mathbf{\alpha}, \mathbf{\Phi_\eta}\right)$ (i.e. multivariate normal distribution with mean zero and variance $\mathbf{\Phi_\eta}$), $\mathbf{\Lambda_Y}$ is the $P \times S$ matrix of factor loadings

between P endogenous indicators y_p and S endogenous latent variables η_s, $\mathbf{v}$ is $Q \times 1$ vector of Q intercepts u_q, and $\boldsymbol{\varepsilon}$ is the $P \times 1$ vector with $MVN(\mathbf{0}, \boldsymbol{\Theta}_{\varepsilon})$ of P observed endogenous unique variables ε_p. The structure model summarizes the relationship between exogenous and endogenous latent variables (i.e. $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$), and can be written as

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

$$(\mathbf{I} - \mathbf{B})\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

$$\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}) \tag{2.5}$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are defined as the above vectors of latent variables, $\boldsymbol{\alpha}$ is the $S \times 1$ vector of S latent factor means (if the factors are not regressed on any predictors) or latent factor intercept (if the factors are regressed on predictors) α_s, $\mathbf{B}$ is the $S \times S$ matrix of S^2 regression coefficients between endogenous latent factors, and $\boldsymbol{\Gamma}$ is the $S \times R$ matrix of SR regression coefficients among endogenous and exogenous latent variables, $\boldsymbol{\zeta}$ is the $S \times 1$ vector with $MVN(\mathbf{0}, \boldsymbol{\Psi})$ of S observed residual ζ_s. Combined with Equation(2.5), Equation(2.2) can be reformatted as

$$\mathbf{y} = \mathbf{v} + \boldsymbol{\Lambda}_{\mathbf{Y}}(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}) + \boldsymbol{\varepsilon} \tag{2.6}$$

Then, the mean and covariance structure of exogenous and endogenous indicators

can be separately stated in the function of unknown model parameters. The mean structure

can be described as

$$\boldsymbol{\mu_X} = E[\mathbf{x}] = \mathbf{u} + \boldsymbol{\Lambda_X} E(\boldsymbol{\xi}) \tag{2.7}$$

$$\boldsymbol{\mu_Y} = E[\mathbf{y}] = \mathbf{v} + \boldsymbol{\Lambda_Y} (\mathbf{I} \text{-} \mathbf{B})^{-1} \left(\boldsymbol{\alpha} + \boldsymbol{\Gamma} E(\boldsymbol{\xi}) \right) \tag{2.8}$$

Define $\mathbf{V}$ as the column vector of vectorization of the observed variables, that is

$\mathbf{V} = [\mathbf{y}, \mathbf{x}] = \left[y_1, y_2, ..., y_p, x_1, x_2, ..., x_q \right]'$. As for the covariance structure (i.e.

$\mathrm{cov}(\mathbf{x}, \mathbf{y}) = E[\mathbf{x}\mathbf{y}'] \text{-} E[\mathbf{x}] E[\mathbf{y}']$), based on the assumptions of a) the orthogonality

between predictors and error and b) uncorrelated error terms between endogenous and

exogenous variables, the variance-covariance matrix can be represented as follows

$$\mathrm{cov}(\mathbf{V}, \mathbf{V}') = \mathrm{cov}([\mathbf{y}, \mathbf{x}], [\mathbf{y}, \mathbf{x}]') = \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \tag{2.9}$$

where

$$\mathrm{cov}(\mathbf{x}, \mathbf{x}) = \mathrm{var}(\mathbf{x}) = \boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \tag{2.10}$$

$$\mathrm{cov}(\mathbf{y}, \mathbf{y}) = \mathrm{var}(\mathbf{y}) = \boldsymbol{\Sigma}_{yy} = \boldsymbol{\Lambda}_y \boldsymbol{\Phi} \boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon \tag{2.11}$$

$$\mathrm{cov}(\mathbf{y}, \mathbf{x}) = \boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta})$$

$$= \mathrm{cov}\left[\left(\mathbf{v} + \boldsymbol{\Lambda_Y}(\mathbf{I}\text{-}\mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}) + \boldsymbol{\varepsilon} \right)\left(\mathbf{u} + \boldsymbol{\Lambda_X}\boldsymbol{\xi} + \boldsymbol{\delta} \right) \right] \tag{2.12}$$

By setting $\boldsymbol{\alpha} = \mathbf{u} = \mathbf{v} = \mathbf{0}$, the variance-covariance matrix of $\mathbf{V}$ can be simplified as

$$\Sigma = \left[\begin{array}{cc} \Sigma_{YY} = \Lambda_Y (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi) \left[(I - B)^{-1} \right]' \Lambda_Y' & \Sigma_{YX} = \Lambda_Y (I - B)^{-1} \Gamma \Phi \Lambda_X' \\ \Sigma_{YX} = \Sigma_{XY} & \Sigma_{XX} = \Lambda_X \Phi \Lambda_X' + \Theta_\delta \end{array} \right]$$

2.1.1 Parameter estimation

Estimation is a procedure to have the estimate of unknown parameter by optimizing

the specific cost function which is composed of hypothesized model and observed data.

Kendall and Stuart (1979) commented "It is impossible, when a parameter space is

between 0 and 1, to construct an unbiased estimator that always takes on values between 0

and 1." Steiger (2000) also commented for point estimation "Seldom, in fact very seldom,

will a statistic be exactly equal to the parameter it is estimating, if the parameter space is

continuous." Despite these doubts, estimation is essentially the most important part in most

of statistical modeling analyses. There are several kinds of estimation methods available,

including Maximum likelihood (ML) estimation, Unweighted Least Squares (ULS)

estimation, Generalized Least Square (GLS) estimation, Weighted Least Squares (WLS)

estimation, Weighted Least Squares Mean-Variance adjustment (WLSMV) estimation,

Asymptotically Distributional Free (ADF) estimation and Maximum a Posterior (MAP)

estimation. Among those estimation methods, Maximum likelihood estimation (MLE) is

the most commonly used ~~one~~ in SEM analysis. MLE is a process to find the estimated

values of unknown parameters by maximizing the likelihood function (i.e. minimizing the

fit function) between the model-implied variance-covariance matrix and observed

variance-covariance matrix. When data fit with certain probabilistic assumptions (e.g. such

as data normality and independency), MLE can not only provide us the MVUE (minimum

variance and unbiased estimate) parameter estimate, but also provide model fit test statistic,

which can be used to assess the quality of hypothesized model, and standard error estimate

of parameter estimate, which can be used to access the quality of parameter estimate.

However, when data exhibit abnormally away from the required probabilistic assumptions,

the test statistic and standard error estimate will lose its ideal properties and result in

erroneous statistical inference conclusion of model fit and parameter estimate. The

problem of using traditional statistical method to deal with non-normal and dependent data

will be discussed in the following section as well as the proposed remedies to resolve the

problem.

Holding with the assumptions of multivariate normality distribution of continuous

observed variables,   residual terms, latent variables, and missing at random (MAR)

pattern of incomplete data, The conditional probability of observed data $\mathbf{y}$ given

variance-covariance matrix $\mathbf{\Sigma}$ and mean vector $\mathbf{\mu}$ can be written as

$$P\left(\mathbf{y}|\mathbf{\Sigma},\mathbf{\mu}\right) = \left(2\pi\right)^{-p/2} \left|\mathbf{\Sigma}\right|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{\mu})'\mathbf{\Sigma}^{-1}\left(\mathbf{y}-\mathbf{\mu}\right)\right) \qquad (2.13)$$

where the $|\mathbf{\Sigma}|$ is the determinant and $\mathbf{\Sigma}^{-1}$ is inverse matrix of $\mathbf{\Sigma}$. The likelihood function

$L(\mathbf{\Sigma}, \mathbf{\mu}; \mathbf{y})$ can be formulated by redefining the Equation(2.13) to be the function of

parameters $\mathbf{\Sigma}$ and $\mathbf{\mu}$, fixing the observed data $\mathbf{y}$; then, the log likelihood function can

be defined by taking the natural logarithm of likelihood function, that is

$l(\mathbf{\Sigma}, \mathbf{\mu}; \mathbf{y}) = \ln L(\mathbf{\Sigma}, \mathbf{\mu}; \mathbf{y})$. By multiplying -2, the fit function of model-implied

variance-covariance matrix $\hat{\mathbf{\Sigma}}(\mathbf{\theta})$ with unknown parameter $\mathbf{\theta}$ and observed/sample

variance-covariance $\mathbf{S}$ can be written as

$$\begin{aligned} F_{ML}\left(\mathbf{S}, \hat{\mathbf{\Sigma}}(\mathbf{\theta})\right) &= \left(\mathbf{V} - \hat{\mathbf{\mu}}(\mathbf{\theta})\right)' \hat{\mathbf{\Sigma}}(\mathbf{\theta})\left(\mathbf{V} - \hat{\mathbf{\mu}}(\mathbf{\theta})\right) \\ &\quad - \ln\left|\hat{\mathbf{\Sigma}}(\mathbf{\theta})\right| + tr\left[\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}(\mathbf{\theta})\right] - \ln|\mathbf{S}| - (P+Q) \end{aligned} \tag{2.14}$$

where $\mathbf{S}$ is the unbiased sample variance-covariance matrix to its population counterpart

$\mathbf{\Sigma}$, $\mathbf{\theta}$ is the $T \times 1$ vector of unknown model parameters, $\mathbf{V}$ is the observed mean vector

and $\hat{\mathbf{\mu}}(\mathbf{\theta})$ is the model-implied mean vector, and $P+Q$ is the number of observed

variables. By minimizing the Equation(2.14), the vector of estimated unknown parameters

$\hat{\mathbf{\theta}}$ can be the consistent and efficient estimates of unknown parameters $\mathbf{\theta}$, that is

$$\hat{\mathbf{\theta}} = \underset{\hat{\mathbf{\theta}}}{\arg\min} F_{ML}\left(\mathbf{S}, \hat{\mathbf{\Sigma}}(\hat{\mathbf{\theta}})\right) = \ln\left|\hat{\mathbf{\Sigma}}(\hat{\mathbf{\theta}})\right| + tr\left|\mathbf{S}\hat{\mathbf{\Sigma}}(\hat{\mathbf{\theta}})\right| - \ln|\mathbf{S}| - (P+Q) \tag{2.15}$$

If the i.i.d. (independent and identical distribution) assumption of independence

and multivariate normality of continuous data holds, the vector of parameter estimates $\hat{\mathbf{\theta}}$

will be the unbiased and efficient estimate (i.e. minimum-variance unbiased estimate,

MVUE) of $\boldsymbol{\theta}$.

2.1.2 Model evaluation

The quality of hypothesized model to the sampled data is evaluated with two

major statistical measures: model-fit test statistics and model-fit fit indices. Most

model-fit fit indices are defined through model-fit test statistics (Yuan, 2005).

Model fit test statistic. When the data normality holds, the test statistics T_{ML}, which is

the product of sample size N and F_{ML} in Equation(2.14), can be used to evaluate the

model fit, i.e.

$$T_{ML} = (N-1) F_{ML}\left(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}\left(\hat{\boldsymbol{\theta}}\right)\right). \tag{2.16}$$

When data are normally distributed and model is correctly specified, T_{ML} will be

close to chi-square distribution χ^2_{df} with the degree of freedom equal $df = \dfrac{r(r+1)}{2} - t$

where r is the number of endogenous and exogenous variables (i.e. $P+Q$) and t is the

number of freely estimated parameters. The hypothesis of whether the hypothesized

model fits the sample data or not can then be tested with chi-square test of exact model fit

$$H_0 : \boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}\left(\hat{\boldsymbol{\theta}}\right)$$
$$H_1 : \boldsymbol{\Sigma} \neq \hat{\boldsymbol{\Sigma}}\left(\hat{\boldsymbol{\theta}}\right)$$

Null hypothesis will be rejected when T_{ML} is larger than the critical value, that is,

hypothesized model does not fit population model. The classical likelihood ratio test

statistic, T_{ML} is a formal model fit evaluation. Several researchers assert this is the only

believable global model-fit indicator, that is, if alternative hypothesis is preferred, the

hypothesized model is not an adequate one to its population counterpart. However, T_{ML}

test statistic is easily influenced by sample size and data non-normality. Just like most

statistical significance tests, researcher can always have significant results for the effect

of any size with sufficiently large sample size. On the other hand, when data is

multivariate normally distributed, the resulted model fit test statistic will behave like the

chi-square distribution. Several studies concluded that T_{ML} is asymptotically robust to

certain data non-normality conditions (Amemiya & Anderson, 1990; Browne & Shapiro,

1988; Hu, Bentler, & Kano, 1992; Kano, 1992; Mooijaart & Bentler, 1991; Satorra, 1992;

Satorra & Bentler, 1990; Yuan & Bentler, 1998, 1999) . However, if the data violate the

normality assumption, the conclusion of statistic inference based on chi-square

distribution will still be erroneous in newer study (Yuan, 2005).

　　　　Model fit index. There are two drawbacks of the interpretation of model fit

likelihood test result. First, the result of model fit hypothesis test only tells us if the

hypothesized model exactly fits the population model or not, that is, if the model-implied

variance-covariance matrix exactly replicates the observed variance-covariance matrix or

not. Second, the model fit test statistic does not have a bounded value, which means you

can have an infinite large value of model fit test statistic but you do not have any idea

whether this value comes from the large sample size of your data, or comes from the

model misspecification of your specified model. As a result, the rejection of null

hypothesis does not tell us the degree of disparity between the population model and

hypothesized model. People are interested in having more information about their

hypothesized model to the sampled data instead of merely a simple right or wrong answer.

Therefore, several model fit indices are developed for this need.

One of the advantages of SEM "is the availability of statistics that assess the

"goodnes of fit" of the hypothesized model" (Campbell-Sills & Brown, 2005, p.22). Model

fit statistics commonly reported in SEM studies include CFI (Comparative Fit Index,

Bentler, 1990), TLI (Tucker-Lewis Coefficient, Bollen, 1989), GFI (Goodness of Fit,

Jöreskog & Sörbom, 1989), RMSEA (Root Means Square Error of Approximation, Steiger

& Lind, 1980; Steiger, 1990), and SRMR (Standardized Root Mean Square Residual,

Bentler, 1995). Among various model fit indices, CFI (relative fit indices) signified how

far the model-implied variance-covariance matrix is away from the null model where there

is no covariance among the observations. In contrast, RMSEA and SRMR (absolute fit

indices) examine how close the model implied variance-covariance matrix based on the

hypothesized model is to the observed variance-covariance matrix based on the observed

data. The formulas of CFI, RMSEA and SRMR can be represented as

$$CFI = 1 - \left[\frac{\max\left(\chi^2_{Model} - df_{Model}, 0\right)}{\max\left(\chi^2_{Null} - df_{Null}, \chi^2_{Model} - df_{Model}\right)} \right] \qquad (2.17)$$

where χ^2_{Model} is the chi-square value with degree of freedom df_{Model} of hypothesized

model, χ^2_{Null} is the chi-square value with degree of freedom df_{Null} of baseline/null model

(e.g. the independent model without any relationship between observed variables);

$$RMSEA = \sqrt{\max\left[\frac{\chi^2_{Model}}{df \times N} - \frac{1}{N}, 0 \right]}; \qquad (2.18)$$

$$SRMR = \sqrt{\frac{\left\{ 2\sum_{i=1}^{P+Q}\sum_{j=1}^{i}\left[\left(s_{ij} - \hat{\sigma}_{ij}\right) / \left(s_{ii}s_{jj}\right)\right]^2 \right\}}{(P+Q)(P+Q+1)}} \qquad (2.19)$$

where N is the sample size, s_{ij} is the observed covariance between variable i with

standard deviation s_{ii} and variable j with standard deviation s_{jj}, and $\hat{\sigma}_{ij}$ is the

model-implied counterpart of s_{ij}. Hu and Bentler (1999) presented a two-index

presentation strategy, in which they suggested researchers at least present two different

kinds of model fit indices to help their readers to assess the quality of their hypothesized

model (such as TLI and SRMR, RMSEA and SRMR, and CFI and SRMR). Hu and Bentler

(1999) also proposed a guidance of the cut-off values for most of model fit indices. Based

on their simulation study, the hypothesized model with CFI > 0.95 and/or RMSEA<0.08

and/or SRMR<0.08 will be thought as the acceptable model to the observed data.

McDonald & Ho (2002) summarized that CFI, RMSEA and SRMR are the most

commonly reported indices in ~~the~~ substanti~~al~~ive research areas using SEM methodology;

Jackson, Gillaspy and Purc-Stephenson (2009) reported that CFI and RMSEA are the most

commonly-used ~~ones~~ in the confirmatory factor analysis (CFA).

2.2 Data Dependency

In social, educational, and organizational research, it is not rare to have data with a

hierarchical or multilevel structure in nature (Heck & Thomas, 2008). In such areas, the

technique of cluster sampling or multistage sampling is widely used due to the efficiency in

time and resources. Unlike simple random sampling (SRS) which randomly select a

sample from a target population to assure that the selected observations are independent

from each other, these complex survey sampling strategies randomly sample naturally

occurring groups/clusters of individuals/observations (Gall, Gall, & Borg, 2006; Stapleton,

2006). In longitudinal studies, researchers keep gathering data from certain participants at

different time points (e.g. the repeated measures of a participant and panel data) also

generate non-independent observations (Duncan, Duncan, & Strycker, 2006; Hardin &

Hilbe, 2002). Conventional statistical methods which assume independent observations

should not be used with data collected from complex survey sampling due to the potential

non-independent observations. The use of conventional statistical methods on

non-independent data can result in biased estimation of the standard errors and incorrect

statistical conclusions (Hox, 2002; Kish, 1995).

Three analytic approaches are usually used for analyzing data collected through

cluster sampling, namely, disaggregated analysis, aggregated analysis, and multilevel

modeling (Hofmann, 1997). Aggregation analysis or disaggregation analysis is a

commonly used methodology when researchers analyze the dependent data. Both methods

specify the analytic model in only a certain level, either in cluster-level (higher-level) or

individual-level (lower-level). The aggregation analysis is conducted using only the

cluster-level information, such as group/cluster mean aggregated from the lower/individual

level, to construct the higher-level analysis. As a result, aggregation analysis will easily fall

into the ecological fallacy (aka. Robinson effect) for the intentionally neglected

lower-level variability (Robinson, 1950). Aggregated data cannot fully reflect the

individual-level variation as well (Au & Cheung, 2004; Klein et al., 2001). Moreover,

studies have shown that regression analysis performed on aggregated data can result in

biased parameter estimates and underestimated standard errors associated with the fixed

effects (Croon & van Veldhoven, 2007; Lüdtke et al., 2008). On the other hand, in

disaggregation analysis, only the lower-level model is constructed and the values of

cluster-level variable are assigned downward to its individual-level counterpart, that is, all

individuals in the same groups will have the same value in the assigned variable from their

higher-level cluster. Consequently, the independence assumption of tradition regression

analysis is violated (Hofmann, 1997; Raudenbush & Bryk, 2002) . Neglecting the

dependency among observations will generally result in underestimating the standard

errors of the fixed effect and lead to inflated Type I error rate (De Leeuw & Kreft, 1995;

Luo & Oi-man Kwok, 2009; Moerbeek, 2004; Raudenbush & Bryk, 2002; Snijders &

Bosker, 1999). It has been shown that none of these two approaches can adequately reveal

the complete picture of the relations between different levels of variables in multilevel data

(Holt et al., 1980; Raudenbush & Bryk, 2002).

Therefore, multilevel analysis is developed to analyze the complex survey data without the above mentioned disadvantages. In general, there are two perspectives in using multilevel modeling to deal with the multilevel data, the natural sampling scheme and the level-varying parameters (Heck & Thomas, 2008; Hox, 2002; Muthén, 1994). For the first perspective, to best fit the actual sampling scheme of multilevel data, the statistical analytic model should be able to capture the random variation of observations in different levels of sampling, such as modeling the participant-level and group-level variance components separately but not the total variance component. In this vein, the analytic model will be constructed according to the actual sampling scheme, such as cluster sampling or multistage sampling. For the second perspective, to acquire the parameter estimates at different level, analytic model should be able to simultaneously calculate the lower-level parameter estimates which can have varying values from different higher-level sampling units.

I presented one example to solidify the concept of why using multilevel modeling is a more appropriate methodology in analyzing dependent data. In general, people in the same group might have similar behavioral pattern sharing but people in different groups might not. When the traditional statistical methods are used, which assume people have no

relationship with one others, the special group characteristics are totally ignored and the

dissimilarity between the different groups of people is averaged out   because we neglect

the homogeneity of participants in the same group and the heterogeneity between groups.

If we conduct the level-specific analyses for different sampling units, such as one analysis

for participants and the other for groups, we will still run into the same problem: the

participant-level analysis ignores the group differences, and the group-level analysis

neglects the individual idiosyncrasy. The best analytical way for the dependent data now is

to simultaneously analyze the models of different levels in one integral multilevel model.

In structural equation modeling (SEM), data are typically assumed to be collected

through SRS so that they are independently and identically distributed (Stapleton, 2006; S.

H. du Toit & M. du Toit, 2008). With this kind of complex survey data, the independent

and identical distributed assumption is violated. Using the conventional SEM modeling

without taking into the consideration of heterogeneous data will result in biased structural

coefficients (Muthén & Satorra, 1989), bias estimation of the standard error of the fixed

effect and erroneous statistical inference of the fixed effect (Hox, 2002; Kaplan & Elliott,

1997b; Kish & Frankel, 1974), and incorrect likelihood ratio test statistics (Muthén &

Satorra, 1995; Yuan, 2005). "By ignoring the hierarchical structure of the data, incorrect

parameter estimates, standard errors, and inappropriate fit statistics may be obtained" (du

Toit & du Toit, 2008, p. 456). In order to have unbiased parameter estimates and consistent

statistical inferences, knowing the hierarchical nature of nested data and specifying this

dependency in analytic model is substantial in various research areas.

In Statistics, there are three major approaches used to account for the extra

correlation in the dependent data, that is scaling, design-based approach with robust

variance estimators, and model-based with hierarchical modeling strategy (Hardin &

Hilbe, 2002). With these three approaches, the violation of i.i.d. assumption of dependent

data is taken into consideration.

Scaling is the easiest method used to adjust standard errors due to perceived

correlation effects. The traditional statistical approach which assumes data from SRS

usually have underestimated standard error estimates. Then, the standard error estimates

of the parameter estimates will be rescaled by dividing by the square root of either the

deviance-dispersion or Pearson χ^2. This kind of scaling approach is a post hoc method

to analyzing dependent data, so it takes no effect on parameter estimates. The major

problem is that scaling only provides an overall adjustment of standard error but does not

capture or adjust for identified clusters or correlation effects.

The design-based approach is using the robust standard error estimator (Huber,

1967; White, 1980) along with the original statistical approach. Sandwich variance

estimator, which is a general name of alternative variance estimators, is another overall

adjustment to the deviated standard error of parameter estimate due to extra dependency.

This kind of relative variance estimator has been proposed to address data

nonindependence (i.e. data heteroskedasticity) more directly. This adjustment is still a

post hoc process and only affects the standard errors but not the parameter estimates.

Under the SEM framework, Muthén and Satorra (1995) proposed using single-level

modeling with ML parameter estimation with robust standard error estimators, such as

Huber-White robust standard error estimator. Besides the adjustment to the standard error,

robust likelihood ratio statistic is also necessary for analyzing the complex survey data.

The model-based approach to account for extra correlation is implementing the

analytic models explicitly specified as the hierarchical structures of heterogeneous data.

Multilevel modeling, also known as Hierarchical linear model, Mixed Effect Model,

Multilevel Regression Model, have been widely investigated and utilized in numerous

areas. Under the SEM framework, this kind of modeling strategy is named as Multilevel

SEM (Goldstein & McDonald, 1988; McDonald & Goldstein, 1989; Muthén, 1989, 1990).

The second and third approaches can be categorized into multilevel modeling strategies which allow researchers to maintain the original hierarchical data structure for analyses. In general, these two methods (design-based and model -based approaches) are commonly used in studies using SEM approaches. Researchers have devoted themselves in modeling the data heterogeneity in the SEM framework. (Aitkin & Longford, 1986; Asparouhov & Muthén, 2005; Boomsma, 1987; Goldstein, 1995; Goldstein, 1987; Goldstein & McDonald, 1988; Hox, 1993; Longford & Muthén, 1992; Mehta & Neale, 2005; Muthén, Khoo, & Gustafsson, 1997; Muthén & Satorra, 1989; Muthén, 1989; Muthén & Asparouhov, 2009; Muthén & Satorra, 1995; Muthén, 1990, 1994; Muthén & Asparouhov, 2002; Satorra & Bentler, 2001; Yuan & Bentler, 1997; Yuan & Hayashi, 2005; Yuan, 2005). For example, for two-level clustered sampling data, the model-based approach analyzes the data by specifying (different) within-level and between-level models respectively whereas the design-based approach analyzes the data with only one overall model and adjusts the standard errors of the parameter estimates based on the sampling design. In this review, we will primarily investigate these two approaches.

2.3 Design-based Approach: Robust Standard Error Estimator and Robust Test Statistics

The design-based approach takes the multilevel data/dependency into account by adjusting for the standard errors of the parameter estimates and the test statistic of model fit hypothesis test based on the sampling design. The design-based approach is commonly used by substantive researchers (Agrawal & Lynskey, 2007; Davidov, Yang-Hansen, Gustafsson, Schmidt, & Bamberg, 2006; Hox & Kleiboer, 2007; Mathews et al., 2009; Muthén & Asparouhov, 2006) given that this approach only requires to specify one single model and that researchers may be mostly interested in only examining the within-level (or level-1) model. For robust standard error estimator, we will start from the normal theory sampling variance estimate (expected Hessian matrix) $\hat{\mathbf{V}}_{\mathbf{EH}}$ and the robust sandwich-type sampling variance estimate $\hat{\mathbf{V}}_{\mathbf{Robust}}$ for Normal Theory Maximum Likelihood (NTML) parameter estimate $\hat{\boldsymbol{\theta}}$ (Hardin & Hilbe, 2007; Huber, 1967; White, 1980), then the robust sandwich-type sampling variance estimate $\tilde{\mathbf{V}}_{Robust}$ for direct ML parameter estimate $\tilde{\boldsymbol{\theta}}$ (Yuan & Bentler, 2000), and finally the robust sandwich-type sampling variance estimate $\breve{\mathbf{V}}_{Robust}$ for pseudo ML parameter estimate $\breve{\boldsymbol{\theta}}$ (Asparouhov & Muthén, 2005). For robust model fit chi-square test statistic, we will first introduce the

Satorra-Bentler rescaled test statistic T_R (Satorra & Bentler, 1988), then Yuan-Bentler

(corrected) ADF test statistic T_{CADF} (Yuan & Bentler, 2000), and finally the pseudo

maximum likelihood test statistic T^* (Asparouhov & Muthén, 2005).

2.3.1 Robust standard error estimator

For statistical inferences, we are interested in knowing not only the location of

population parameter estimates but also the accuracy of parameter estimates. The

sampling variance, whose square rooted values is the so called standard error, is used as

the measure of quality of parameter estimates. The smaller the sampling variance and

standard error, the smaller the fluctuation of the parameter estimates from the repeatedly

sampled samples. However, this does not mean that the smaller the standard error the

closer the parameter estimate to its population location (even the experienced researcher

will sometimes give the erroneous statement like this). If you want to draw this

conclusion, the first thing you need to make sure is that the parameter estimate is the

unbiased estimate of its true population value. Take the terminology of measurement

theory for example, sampling variance is the measure of how reliable your parameter

estimate is, and should not be inferred as the measure of validity. If the parameter

estimate is not a valid indicator of the location of true population value at any degree (i.e.

there is a bias between parameter estimate and its true value), by knowing its smaller

sampling variance, we can only conclude this parameter estimate is a reliable but not a

valid measure of population true value.

It has been known that NTML estimation, under the multivariate normality

assumption, still produces unbiased parameter estimates (i.e. the location of parameter

estimate is consistent to its population location) when the empirical data are actually not

normally distributed (Longford, 1993; Muthén & Asparouhov, 2002). However, the

quality of parameter estimate, that is, the standard error, is seriously influenced by the

sampling design (Kish, 1995; Stapleton, 2008). We will have biased standard error

estimate of the fixed effect and erroneous statistical inference of the fixed effect when we

ignore the extra data dependency (Hox, 2002; D. Kaplan & Elliott, 1997; Kish & Frankel,

1974). In linear regression framework, it has been shown that the standard error of the

regression coefficient for variable from the neglected higher-level is underestimated (Luo

& Oi-man Kwok, 2009; Moerbeek, 2004). Muthén and Asparouhov (2002) also mentioned

this phenomenon when NTML estimation is used to analyze non-normal data in the

framework of latent variable model.

In reality, we can hardly conduct numbers of repeated samplings/experiments to

have the empirical estimate of sampling variance (square of standard error). So in ML

estimation framework, a mathematical measure which can be used to be the estimate of

sampling variance estimate was developed, i.e. the Hessian of the likelihood function of

the one-time sampled data. Assuming $P\left(\mathbf{x}_g\middle|\mathbf{\theta}\right)$ is the conditional probability of observing

the subset of overall survey samples $\mathbf{x}_g$, $\forall g = 1, 2, ..., G$, given the population parameter

$\mathbf{\theta}$. By putting the sample subsets into matrix, that is, overall observed sample matrix

$\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_g, ..., \mathbf{x}_G\right]$, the likelihood of population parameter $\mathbf{\theta}$ given observed

sample matrix $\mathbf{X}$ can be denoted as $L\left(\mathbf{X};\mathbf{\theta}\right)$. Although conditional probability and

likelihood are at two sides of a mathematic formula, they represent very different

meanings in Statistics. Then, in estimation theory, the natural logarithm function of

likelihood, $l\left(\mathbf{X};\mathbf{\theta}\right) = \ln L\left(\mathbf{X};\mathbf{\theta}\right)$, is often used due to the simplicity of computation of the

probability density function (*pdf*) with the exponential power terms from the parametric

exponential distribution family (i.e. Bernoulli, Binomial, Poisson, and Gaussian, etc.) (e.g.

$\ln\left(ab\right) = \ln a + \ln b$, and $\ln\left(a^b\right) = b \ln a$). The score (i.e.

$$\frac{\partial}{\partial \mathbf{\theta}} \ln L\left(\mathbf{X};\mathbf{\theta}\right) = \frac{1}{L\left(\mathbf{X};\mathbf{\theta}\right)} \frac{\partial L\left(\mathbf{X};\mathbf{\theta}\right)}{\partial \mathbf{\theta}}$$, the partial derivative of log likelihood w.r.t. $\mathbf{\theta}$,) will

be set to equal zero to have the maximum likelihood parameter estimate $\hat{\mathbf{\theta}}$ of

population parameter $\boldsymbol{\theta}$. Score is also the sensitivity of log likelihood (i.e. gradient).

After we have the parameter estimate, we will use the higher moment of log likelihood

function to have the estimate of standard error of parameter estimate. So, by setting the

regularity condition of score, (i.e. $E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \middle| \boldsymbol{\theta}\right] = 0$), the variance of score (aka.

the well-known Fisher information or Expected information) can be formulated as

$$
\begin{aligned}
\mathbf{I}(\boldsymbol{\theta}) &= E\left[\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta})\right]^2 \middle| \boldsymbol{\theta}\right] = Var\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta})\right] \\
&= E\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln L(\mathbf{X};\boldsymbol{\theta}) \middle| \boldsymbol{\theta}\right]
\end{aligned}
\tag{2.20}
$$

In fact, it has been shown that parameter estimate $\hat{\boldsymbol{\theta}}$ is an asymptotically

normally distributed and consistent estimate to the population parameter $\boldsymbol{\theta}$ (Yuan &

Jennrich, 1998). Sir R. A. Fisher gave a formal definition of the information: the

reciprocal of the square of standard error (i.e. sampling variance) of estimating a

parameter, i.e.

$$
I(\boldsymbol{\theta}) = \frac{1}{\mathbf{SE}^2(\boldsymbol{\theta})} = \frac{1}{\hat{\mathbf{V}}_{\mathbf{EH}}(\boldsymbol{\theta})}
\tag{2.21}
$$

where $\hat{V}_{EH}(\boldsymbol{\theta})$ is the expected Hessian matrix used to be the estimate of sampling

variance of parameter estimate (Hardin & Hilbe, 2007). Given the likelihood function

$L(\mathbf{X};\boldsymbol{\theta})$, it is a way of measuring the amount of information that an observed random

matrix $\mathbf{X}$ carries about an unknown population parameter $\boldsymbol{\theta}$. It is named as

information because it tells us how much information is in the score about $\boldsymbol{\theta}$ (e.g. the

larger the information, the smaller the sampling variance of unknown parameter). Same

idea is also convoyed in Item Response Theory.

The above mentioned information statistics is a consistent and efficient estimator

of the standard error of parameter estimate $\hat{\boldsymbol{\theta}}$ when data is normally distributed, that is,

the Hessian matrix is a valid and minimum asymptotic covariance estimator of sampling

variance when the normality assumption is met and the specification of the variance in

Equation(2.20) is correct. However, for real data, the normality assumption is almost

always violated. To solve this abnormality, the sandwich-type estimator (Hardin & Hilbe,

2007; Huber, 1967; White, 1980) can produce consistent estimate of sampling variance

with the formula for calculating the variance components including a score factor

"sandwiched" between two copies of the Hessian matrix, that is,

$$\hat{\mathbf{V}}_{\mathbf{Robust}} = \hat{\mathbf{V}}_{\mathbf{EH}} \hat{\mathbf{V}}_{\mathbf{OPG}} \hat{\mathbf{V}}_{\mathbf{EH}} \qquad (2.22)$$

where $\hat{\mathbf{V}}_{\mathbf{OPG}}$ is the outer product of two gradient functions. Here we define gradient of

log likelihood function as its first derivative but without the expectation, i.e.

$\dfrac{\partial}{\partial \theta} \ln L(\mathbf{X}; \boldsymbol{\theta})$. So, Equation(2.22) can be written as

$$\hat{\mathbf{V}}_{\mathbf{Robust}} = \hat{\mathbf{V}}_{\mathbf{EH}} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \times \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \right\} \hat{\mathbf{V}}_{\mathbf{EH}} \tag{2.23}$$

$$= \mathbf{I}^{\mathbf{-1}} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \times \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \right\} \mathbf{I}^{\mathbf{-1}} \tag{2.24}$$

$$= \left\{ E\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln L(\mathbf{X};\boldsymbol{\theta}) \middle| \boldsymbol{\theta} \right] \right\}^{-1} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \times \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{X};\boldsymbol{\theta}) \right\} \left\{ E\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln L(\mathbf{X};\boldsymbol{\theta}) \middle| \boldsymbol{\theta} \right] \right\}^{-1} \tag{2.25}$$

When data are normally distributed and their missingness pattern is based on at least missing at random (MAR), $\hat{\mathbf{V}}_{\mathbf{OPG}}$ is a sufficient estimator to the normal theory information matrix (i.e. $\hat{\mathbf{V}}_{\mathbf{OPG}} = \mathbf{I} = \hat{\mathbf{V}}_{\mathbf{EH}}^{\mathbf{-1}}$), and the Hessian matrix is again the consistent and efficient estimator to the sampling variance and the sandwich estimate of sampling variance is equal to the normal theory Hessian matrix (i.e. $\hat{\mathbf{V}}_{\mathbf{Robust}} = \hat{\mathbf{V}}_{\mathbf{EH}}$) (Hardin & Hilbe, 2007; Yuan & Bentler, 2000).

Huber (1967) first presented this sandwich type variance estimator in some weak non-normality conditions; White (1980) independently gave this robust variance estimate in heteroskedasticity condition under linear model framework. In order to credit their works, this sandwich type robust standard error estimator is also called Huber-White robust standard error estimator (aka. survey variance estimator, design-based variance estimator, and empirical variance estimator) and has been widely used in survey statistics (Hardin & Hilbe, 2007; Kish, 1995). With the use of this robust estimator, an asymptotically consistent estimate of covariance matrix can be derived free from

distributional assumptions of observations (Hardin & Hilbe, 2007; Huber, 1967; White,

1980). Readers can refer to Hardin and Hilbe (2002, 2007) for readable descriptions of

varieties of sandwich-type variance estimators. Diggle, Liang, and Zeger (1994) stated

that sandwich-type robust standard error estimator is best used when data come from

"many experimental units" in their longitudinal research book; Muthén and Satorra (1995)

also concluded that this kind of robust standard error estimator is useful in dealing with

the extra dependency in complex survey data in their simulation study.

Except for ML estimation, the direct ML estimation method (Yuan & Bentler,

2000) (aks. Full information maximum likelihood estimation, FIML/FML (Arbuckle,

1996; Enders & Bandalos, 2001)) is another feasible estimation method to calculate the

unknown parameter estimate $\tilde{\boldsymbol{\theta}}$. The direct ML estimation is commonly used in dealing

with the data non-normality issue of incomplete data in latent variable model and

Structural equation modeling (Allison, 1987; Arbuckle, 1996; Finkbeiner, 1979; Enders,

2008; Graham, 2003; Lee, 1986; Lee & Song, 2007; Muthén, Kaplan, & Hollis, 1987;

Yuan & Bentler, 2000). Muthén and Satorra (1995) mentioned that the reason why the

observed data is usually not normally distributed is because that the survey sample is just

a small portion of target population, and the resulted observed variables are often skewed.

So, the techniques used in analyzing the nonnormal data should be essentially special

cases of those in dealing with complex survey data. Accordingly, the robust standard

error estimator from direct ML estimation method is used in SEM framework with

dependent data to remedy the underestimated SE estimate from NTML estimation

(Muthén & Satorra, 1995; Muthén & Muthén, 1998; Stapleton, 2006).

In direct ML estimation, the summation of case-wise likelihood functions of the

subsets (e.g. $\mathbf{x}_g$) of overall data (e.g. $\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_g, ..., \mathbf{x}_G \right]$) will be maximized to

have the vector of unknown parameter estimates $\tilde{\boldsymbol{\theta}}$. The individual case-wise log

likelihood function $l\left(\mathbf{x}_g; \boldsymbol{\theta}\right)$ of subset of observed data $\mathbf{x}_g$ is defined as

$$l\left(\mathbf{x}_g; \boldsymbol{\theta}\right) = \ln L\left(\mathbf{x}_g; \boldsymbol{\theta}\right)$$

$$= \ln\left(2\pi\right)\frac{rank\left(\mathbf{V}_g\right)}{2} - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_g\left(\boldsymbol{\theta}\right)\right| + \left(\bar{\mathbf{V}}_g - \boldsymbol{\mu}_g\left(\boldsymbol{\theta}\right)\right)' \boldsymbol{\Sigma}_g\left(\boldsymbol{\theta}\right)\left(\bar{\mathbf{V}}_g - \boldsymbol{\mu}_g\left(\boldsymbol{\theta}\right)\right) \qquad (2.26)$$

where $\mathbf{V}_g$ is the vectorization of observed variables $\mathbf{V}_g = \left[\mathbf{y}_g; \mathbf{x}_g\right]'$, $rank\left(\mathbf{V}_g\right)$ is the

number (i.e. $P_g + Q_g$) and $\bar{\mathbf{V}}_g$ is the mean vector of the observed variables (i.e.

$\bar{\mathbf{V}}_g = E\left[\mathbf{V}_g\right]$), and the mean vector $\boldsymbol{\mu}_g\left(\boldsymbol{\theta}\right)$ and variance-covariance matrix $\boldsymbol{\Sigma}_g\left(\boldsymbol{\theta}\right)$ are

the subset of overall mean vector $\boldsymbol{\mu}\left(\boldsymbol{\theta}\right)$ and variance-covariance matrix $\boldsymbol{\Sigma}\left(\theta\right)$. Then,

the direct log-likelihood function is the summation of the case-wise log-likelihood

functions across entire sample, that is,

$$l(\mathbf{x};\boldsymbol{\theta}) = \sum_{g=1}^{G} l(\mathbf{x}_g;\boldsymbol{\theta}) \qquad (2.27)$$

Then by maximizing the Equation(2.27), the vector of estimated unknown

parameters $\tilde{\boldsymbol{\theta}}$ can be the consistent and efficient estimates of unknown parameters $\boldsymbol{\theta}$,

that is

$$\tilde{\boldsymbol{\theta}} = \underset{\tilde{\boldsymbol{\theta}}}{\arg\max}\, l(\mathbf{x};\boldsymbol{\theta}) \qquad (2.28)$$

If we have complete and continuous dataset holding the i.i.d. assumption of

independence and multivariate normality, the direct ML estimate $\tilde{\boldsymbol{\theta}}$ will be identical to

the traditional NTML estimate $\hat{\boldsymbol{\theta}}$ from Equation(2.15). Then the similar robust standard

error estimator as Equation(2.22) can be constructed by using the arithmetic mean

($\bar{\mathbf{X}} = \dfrac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$) to replace the expectation ($E[\mathbf{X}]$), and the robust standard error estimate

(Arbuckle, 1996; Muthén & Asparouhov, 2002; Yuan & Bentler, 2000) of direct ML

estimate can be formulated as

$$\tilde{\mathbf{V}}_{Robust} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} \qquad (2.29)$$

with $\mathbf{A} = \dfrac{1}{G}\sum_{g=1}^{G} -\dfrac{\partial^2}{\partial\boldsymbol{\theta}^2} l(\mathbf{x}_g;\boldsymbol{\theta})$, $\mathbf{B} = \left(\dfrac{1}{G}\sum_{g=1}^{G} \dfrac{\partial}{\partial\boldsymbol{\theta}} l(\mathbf{x}_g;\boldsymbol{\theta}) \times \dfrac{\partial}{\partial\boldsymbol{\theta}} l(\mathbf{x}_g;\boldsymbol{\theta})\right)$.

Asparouhov and Muthén (2005) discussed the same robust sandwich standard

error estimator as Equation(2.29) using the Pseudo Maximum Likelihood (PML)

estimation (Skinner, 1989), which is essentially the same as direct maximum likelihood

function but with sampling weight in it, that is

$$l(\mathbf{x};\boldsymbol{\theta}) = \sum_{g=1}^{G} w_g l(\mathbf{x}_g;\boldsymbol{\theta}) \tag{2.30}$$

where $w_g = 1/p_g$ and p_g denotes the sampling probability of participant g who is

sampled in the survey sample. The unknown parameter estimate $\breve{\boldsymbol{\theta}}$ is calculated by

maximizing the concave PML function, that is, set the first derivative of Equation(2.30)

equal zero and solve this equation

$$\breve{\boldsymbol{\theta}} = \underset{\breve{\boldsymbol{\theta}}}{\arg\max}\, l(\mathbf{x};\boldsymbol{\theta}) \tag{2.31}$$

Then, the robust sandwich-type standard error estimate of $\breve{\boldsymbol{\theta}}$ can be derived as

Equation(2.29). This kind of robust sandwich standard estimate has been implemented in

Type=Complex and Estimator=MLR routine in Mplus (Muthén & Muthén, 1998).

2.3.2 Robust model fit test statistics

Besides underestimated standard error estimate, ignoring modeling extra data

dependency/non-noramlity will also result in incorrect model fit likelihood ratio test

statistics (Muthén & Satorra, 1995; Satorra & Bentler, 2001; Yuan & Bentler, 1997; Yuan,

2005). Although several simulation studies (see the above cited references) showed that

T_{ML} in Equation(2.16) is asymptotically robust to certain data non-normality conditions,

Yuan (2005) warned the researchers against "blindly" trusting that T_{ML} will "really"

asymptotically close to χ^2_{df} distribution with non-normally distributed data before the

formal definition and procedure of asymptotic robustness is settled down. Three

commonly used rescaled chi-square test statistics for mean and covariance structure are

discussed here: Satorra-Bentler rescaled normal theory test statistic (Satorra & Bentler,

1988), Yuan-Bentler corrected ADF test statistic (Yuan & Bentler, 2000), and pseudo

maximum likelihood test statistic (Asparouhov & Muthén, 2005).

Satorra and Bentler (1988) presented a rescaled normal theory chi-square test

statistic, named Satorra-Bentler rescaled test statistic, which is an adjusted normal theory

chi-square test statistic which penalizes the normal theory chi-square for the amount of

kurtosis in the data.

$$T_R = \hat{\kappa}^{-1} T_{ML} \qquad (2.32)$$

where $\hat{\kappa}$ is the common kurtosis estimate. Two kinds of opinions differ on the

performance of T_R: some simulation studies suggest T_R performs robustly on some data

non-normality conditions (Chou, Bentler, & Satorra, 1991; Hu et al., 1992); others query

these simulation studies about their unclearness of data generation and conclude T_R does

not approach chi-square distribution and may result in erroneous statistical inference

(Yuan & Bentler, 1999; Yuan, 2005). The Satorra-Bentler rescaled chi-square test

statistic is implemented in various SEM softwares, such as AMOS (Arbuckle, 2003),

LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 1995), and Mplus (Estimator =

MLM) (Muthén & Muthén, 1998).

Browne (1984) proposed asymptotically distribution free (ADF) statistic T_{ADF} : as

long as the data have finite 4th moment (i.e. finite kurtosis measure), T_{ADF} will

asymptotically approaches χ^2_{df} . However, the mean and variance of the T_{ADF}

distribution will be larger than those of χ^2_{df} distribution for practical sample sizes (i.e. the

desired ADF property can be achieved only when sample size goes to infinity) (Hu et al.,

1992). So, we will have inflated type I error rate to reject the correctly specified models if

we still use T_{ADF} test statistic with χ^2_{df} distribution to draw the statistical inference. In

order to have a better performance statistics with smaller sample size, Yuan and Bentler

(1997) proposed a corrected T_{CADF} statistic. Yuan (2005) concluded that the average of

T_{CADF} from sufficient numbers of simulations will be closed to the degree of freedom of

hypothesized model for most sample sizes across various distributions of observed

variables. However, in small sample sizes data conditions, rejection rate of T_{CADF} with

correctly specified models is smaller than the nominal level (e.g. p-value =0.05), and

there is still the non-convergence problem because of using ADF estimation method.

By the same logic of above two robust model fit test statistic, Asparouhov and

Muthén (2005) described an adjusted likelihood ratio test (LRT) statistic with pseudo

likelihood shown in Equation(2.30). In general, the distribution of the LRT test statistic

coming from maximizing the pseudo/weighted log-likelihood function will not approach

χ^2_{df} , and the distribution will be seriously influenced by the sampling design. So,

Asparouhov and Muthén (2005) presented the adjusted LRT statistic taking the sampling

design into account using the pseudo likelihood functions as

$$T^* = c \cdot 2 \left(L_1 - L_2 \right) \tag{2.33}$$

$$\text{with} \quad c = \frac{d_1 - d_2}{trace\left[\mathbf{A}_1^{-1}\mathbf{B}_1 \right] - trace\left[\mathbf{A}_2^{-1}\mathbf{B}_2 \right]} \text{, and}$$

$$\mathbf{A}_i = \frac{1}{G}\sum_{g=1}^{G} -\frac{\partial^2}{\partial\mathbf{\theta}_i^2} l\left(\mathbf{x}_g ; \mathbf{\theta}_i \right), \text{ and}$$

$$\mathbf{B}_i = \frac{1}{G}\sum_{g=1}^{G} \frac{\partial}{\partial\mathbf{\theta}_i} l\left(\mathbf{x}_g ; \mathbf{\theta}_i \right) \times \frac{\partial}{\partial\mathbf{\theta}_i} l\left(\mathbf{x}_g ; \mathbf{\theta}_i \right)$$

where the subscript $i = 1, 2$ indicates two nested model M_1 and M_2, L_i is the

respective pseudo likelihood functions, and d_i denotes the respective numbers of free

parameters and $d_1 - d_2$ is the degree of freedom for the hypothesized model fit test.

According to their simulation study (Asparouhov & Muthén, 2005), T^* is approximately

close to $\chi^2_{d_1-d_2}$ distribution. One benefit of this adjusted test statistic is the ability of

dealing with the incomplete data due to the use of sampling weighted direct-ML-like

estimation method (i.e. Pseudo ML estimation). However, we still do not have a clear

picture of the performance of T^* across various conditions. This adjusted chi-square test

statistic has been implemented in Mplus (Estimator = MLR) (Muthén & Muthén, 1998),

which is the advanced version of Estimator =MLM with the ability to dealing the

incomplete dataset.

2.4 Model-based Approach: Multilevel SEM

There are various kinds for multilevel modeling methodology: Mixed Model

(Henderson, 1975; Littell, Milliken, Stroup, & Wolfinger, 1996; Littell, Milliken, Stroup,

Wolfinger, & Schabenberber, 2006), unified Random-effect model approach for repeated

measures/panel data (Laird & Ware, 1982), random parameter model for dependent data

(Aitkin & Longford, 1986), Variance component model (Anderson & Aitkin, 1985;

Longford, 1987), Random coefficient model (Longford, 1993), Hierarchical Linear

Model (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002), Multilevel Regression

Model (Hox, 2002), contextual modeling (Bovaird, 2007; Kreft & De Leeuw, 1998) and

Multilevel SEM model (Heck & Thomas, 2008; Hox, 2002; Kaplan, 2008; Mehta & Neale, 2005; Muthén, 1994).

In this decade, researchers in organizational, educational and psychological fields started to apply the model-based SEM techniques (Branum-Martin et al., 2006; Cheung & Au, 2005; Duncan, Alpert, & Duncan, 1998; Dyer, Hanges, & Hall, 2005; Everson & Millsap, 2004) given that the model-based approach allows us not only to analyze hierarchical data simultaneously by specifying both within- and between-level models, but also to consider the measurement errors in constructing the latent variables in different levels (Heck & Thomas, 2008; Hox, 2002; David W. Kaplan, 2008; Rabe-Hesketh & Skrondal, 2008; du Toit & du Toit, 2008). Several SEM softwares has provided the multilevel modeling routines for analyzing complex survey data, such as EQS 6, LISREL 8.5 (Jöreskog & Sörbom, 1996) and Mplus (Muthén & Muthén, 1998).Those interested in the details of multilevel SEM (MSEM) methodology can refer to these nice review papers and book chapters (Bovaird, 2007; Curran, 2003; Heck & Thomas, 2008; Kaplan & Elliott, 1997b; Kaplan, 2008; Muthén, 1994; Rabe-Hesketh, Skrondal, & Zheng, 2007; du Toit & du Toit, 2008).

Take a multilevel data drawn from a two-level multistage sampling strategy as an

example. Suppose that G groups are randomly drawn from the target population at the

first stage of sampling, and then n_g participants are sampled within each group g at the

second stage. We have total $N = \sum_{g=1}^{G} n_g$ participants. For each participant, P response

variables are gathered. We now have N $1 \times P$-dimensional random vector $\mathbf{y}_{ig}$ for

participant i (level 1 unit) within group g (level 2 unit) with the P elements y_{pig},

$p = 1, 2, ..., P$,

$$\mathbf{y}_{ig} = \left[y_{1ig}, y_{2ig}, ..., y_{Pig} \right] \tag{2.34}$$

Thus, for each gth group the random matrix of observations can be arranged as:

$$\mathbf{y}_g = \begin{bmatrix} \mathbf{y}_{1g} \\ \mathbf{y}_{2g} \\ \vdots \\ \mathbf{y}_{Ig} \end{bmatrix} = \begin{bmatrix} \left[y_{11g} & y_{21g} & \cdots & y_{P1g} \right] \\ \left[y_{12g} & y_{22g} & \cdots & y_{P2g} \right] \\ \vdots & \vdots & \cdots & \vdots \\ \left[y_{1Ig} & y_{2Ig} & \cdots & y_{PIg} \right] \end{bmatrix} \tag{2.35}$$

Analogous to the variance decomposition used in ANOVA analysis, the

observation $\mathbf{y}_{ig}$ in Equation (2.34) can be decomposed into its between-group component

and within-group component, that is,

$$\mathbf{y}_{ig} = \mathbf{y}_{B..g} + \mathbf{y}_{W.ig}, \quad \forall i = 1, 2, ..., I; \quad g = 1, 2, ..., G \tag{2.36}$$

where $\mathbf{y}_{B..g}$ is the between-group component with $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma_B})$ (i.e. multivariate

normality distribution with grand mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma_B}$) and

$\mathbf{y}_{W.ig}$ is the within-group component with $MVN(\boldsymbol{\mu}_g, \boldsymbol{\Sigma_W})$(i.e. multivariate normality

distribution with grand mean $\boldsymbol{\mu}_g$ and variance-covariance matrix $\boldsymbol{\Sigma_B}$). The correlation

between between-group component in different groups is set to be uncorrelated, that is, the

$Cov(\mathbf{y}_{B..g}, \mathbf{y}_{B..g'}) = \mathbf{0}, \ \forall g \neq g'$. In the same vein, the correlation between different

participants within the same group or in different groups is also set to be zero (i.e.

$Cov(\mathbf{y}_{W.ig}, \mathbf{y}_{W.i'g}) = \mathbf{0}, \ \forall i \neq i'$; $Cov(\mathbf{y}_{W.ig}, \mathbf{y}_{W.i'g'}) = \mathbf{0}, \ \forall i \neq i' \ \& \ \forall g \neq g'$). Furthermore,

the cross-level correlation between $\mathbf{y}_g^B$ and $\mathbf{y}_{W.ig}$ is defined as uncorrelated, that is,

$$Cov(\mathbf{y}_{B..g}, \mathbf{y}_{W.ig}) = \mathbf{0}, \ \forall i = 1, 2, ..., I ; \ g = 1, 2, ..., G$$

Hence, the variance-covariance matrix of $\mathbf{y}_{ig}$ can be decomposed into the

combination of between-group and within-group variations,

$$Cov(\mathbf{y}_{ig}) = \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W \tag{2.37}$$

To take a step further to consider the latent variable model, the Equation (2.36) can

be written as

$$\begin{aligned} \mathbf{y}_{ig} &= \mathbf{y}_{B..g} + \mathbf{y}_{W.ig} \\ &= \boldsymbol{\mu}_B + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B..g} + \boldsymbol{\varepsilon}_{B..g} + \boldsymbol{\mu}_W + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{W.ig} + \boldsymbol{\varepsilon}_{W.ig} \end{aligned} \tag{2.38}$$

The between-group component $\mathbf{y}_{B..g}$ is the combination of intercept vector $\boldsymbol{\mu}_B$, product of factor loading matrix $\boldsymbol{\Lambda}_B$ and latent factor $\boldsymbol{\eta}_{B..g} \sim MVN(\mathbf{0}, \boldsymbol{\Psi}_B)$, and the unique vector $\boldsymbol{\varepsilon}_{B..g} \sim MVN(\mathbf{0}, \boldsymbol{\Theta}_B)$. The within-group component $\mathbf{y}_{W.ig}$ is the combination of intercept vector $\boldsymbol{\mu}_W$, product of factor loading matrix $\boldsymbol{\Lambda}_W$ and latent factor $\boldsymbol{\eta}_{W.ig} \sim MVN(\mathbf{0}, \boldsymbol{\Psi}_W)$, and the unique vector $\boldsymbol{\varepsilon}_{W.ig} \sim MVN(\mathbf{0}, \boldsymbol{\Theta}_W)$. Additional assumption of the relationships among random components is

$$\boldsymbol{\eta}_{B..g} \perp \boldsymbol{\varepsilon}_{B..g} \perp \boldsymbol{\eta}_{W.ig} \perp \boldsymbol{\varepsilon}_{W.ig} \tag{2.39}$$

Different from Equation (2.2), Equation (2.38) specifies two sources of random variation for the observed variables, within group variation and between group variation, rather than just one random source. So, the variance-covariance matrix of $\mathbf{y}_{ig}$ in Equation (2.37) can be further rewritten as

$$\begin{aligned}
Cov(\mathbf{y}_{ig}) &= \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W \\
&= Cov(\boldsymbol{\mu}_B + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B..g} + \boldsymbol{\varepsilon}_{B..g} + \boldsymbol{\mu}_W + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{W.ig} + \boldsymbol{\varepsilon}_{W.ig}) \\
&= Cov(\boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B..g} + \boldsymbol{\varepsilon}_{B..g}) + Cov(\boldsymbol{\Lambda}_W \boldsymbol{\eta}_{W.ig} + \boldsymbol{\varepsilon}_{W.ig}) \\
&= \boldsymbol{\Lambda}_B \boldsymbol{\Psi}_B \boldsymbol{\Lambda}_B' + \boldsymbol{\Theta}_B + \boldsymbol{\Lambda}_W \boldsymbol{\Psi}_W \boldsymbol{\Lambda}_W' + \boldsymbol{\Theta}_W
\end{aligned} \tag{2.40}$$

An important measure of heterogeneity of clusters of dependent data, the intra-class correlation (ICC) is defined as the ratio between cluster-level variance and the total variance of a variable (Cohen, Cohen, West, & Aiken, 2003; Muthén & Satorra, 1995;

Shrout & Fleiss, 1979). The indicator variance is a function of random effects and fixed

effects, no matter in between- and within-level models. Take the multilevel CFA model for

example. According to Equation(2.40), the indicator variance can be expressed as a

combination of three components: factor loadings between indicators and latent factors,

latent factor variances (explained portion of indicator variance), and residual variance of

indicators (unexplained portion of indicator variance). Therefore, the ICC of CFA model

can be controlled by adjusted portions of one of these three components in between- and

within-levels, holding the other two components constant across two levels. For instance,

ICC of multilevel CFA model can then be expressed as the ratio between the between-level

latent factor variance and total latent factor variance, with factor loadings and residual

variance being fixed (Muthén, 1991, 1994), that is,

$$ICC = \frac{\Psi_B}{\Psi_B + \Psi_W}$$

Snijders and Bosker (1999) suggested that the ICC value should be more than

0.05 to construct the multilevel analysis. Hox and Maas (2001) mentioned that in

educational research, the ICCs are often less than 0.20, but in family research or in social

network analysis with sociometric status, the ICCs sometimes will exceed 0.33.

2.4.1 Parameter estimation

Beginning from the late 80s', researchers started to devote themselves in developing the suitable multilevel SEM model with respective estimation strategy for the dependent data (Goldstein, 1987; Goldstein & McDonald, 1988; McDonald & Goldstein, 1989; Mehta & Neale, 2005; Muthén & Satorra, 1989; Muthén, 1989; Muthén & Satorra, 1995; Muthén, 1994). From Equation(2.36), we further decompose the total score into the between -group component and within-group component (Cronbach & Webb, 1975), that is

$$
\begin{aligned}
\mathbf{y}_{ig} &= \mathbf{y}_{B..g} + \mathbf{y}_{W.ig} \\
&= \bar{\mathbf{y}}_g + \left(\mathbf{y}_{ig} - \bar{\mathbf{y}}_g \right)
\end{aligned}
\tag{2.41}
$$

Then, with the assumption of Equation(2.39), the above between-/within-group scores are orthogonal and additive to each other (Hox & Maas, 2004; Muthén & Satorra, 1995; Muthén, 1994). With the score decomposition in Equation(2.41), the independent assumption of different level latent factors shown in Equation(2.39) can be established because the lower-level group-centered scores are uncorrelated with the group means. We then construct the population covariance matrix and the between/within covariance components as

$$
\boldsymbol{\Sigma}_{\mathbf{T}} = Cov\left(\mathbf{y}_{ig}\right) = Cov\left(\bar{\mathbf{y}}_g\right) + Cov\left(\left(\mathbf{y}_{ig} - \bar{\mathbf{y}}_g \right)\right) = \boldsymbol{\Sigma}_{\mathbf{B}} + \boldsymbol{\Sigma}_{\mathbf{W}}
\tag{2.42}
$$

The corresponding between- and within-covariance components will be orthogonal and additive (Muthén, 1994; Searle, Casella, & McCulloch, 1992).

Same score decomposition can be performed for the observed sample data, and the resulted variance-covariance matrix can be shown as

$$\mathbf{S_T} = \mathbf{S_B} + \mathbf{S_W} \tag{2.43}$$

where $\mathbf{S_B}$ and $\mathbf{S_W}$ are the consistent and unbiased estimators to their population counterparts, $\mathbf{\Sigma_B}$ and $\mathbf{\Sigma_W}$, respectively (Heck & Thomas, 2008; Hox, 2002; Hox & Maas, 2004; Muthén, 1994).

With this idea of variance-covariance matrix decomposition, Muthén (1989, 1990, 1994) presented an a partial Maximum likelihood estimation method, also named MUML (Muthén's ML). In MUML, two variance-covariance matrices of different levels are constructed as

$$\mathbf{S_T} = \mathbf{S_{B,MUML}} + \mathbf{S_{PW,MUML}} \tag{2.44}$$

The above three variance-covariance matrices are defined as

$$\mathbf{S_T} = \frac{1}{N-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} \left(\mathbf{y}_{gi} - \bar{\mathbf{y}} \right) \left(\mathbf{y}_{gi} - \bar{\mathbf{y}} \right)'$$

$$\mathbf{S_{PW,MUML}} = \frac{1}{N-G} \sum_{g=1}^{G} \sum_{i=1}^{N_g} \left(\mathbf{y}_{gi} - \bar{\mathbf{y}}_g \right) \left(\mathbf{y}_{gi} - \bar{\mathbf{y}}_g \right)'$$

$$\mathbf{S}_{\mathbf{B,MUML}} = \frac{1}{G-1} \sum_{g=1}^{G} \left(\overline{\mathbf{y}}_g - \overline{\mathbf{y}} \right) \left(\overline{\mathbf{y}}_g - \overline{\mathbf{y}} \right)'$$

where the pooled within-level observed variance-covariance matrix $\mathbf{S}_{\mathbf{PW,MUML}}$ is the

consistent and unbiased estimator to $\mathbf{\Sigma}_{\mathbf{W}}$, and the scaled between-level observed

variance-covariance matrix $\mathbf{S}_{\mathbf{B,MUML}}$ is the consistent and unbiased estimator to $\mathbf{\Sigma}_{\mathbf{W}} + c\mathbf{\Sigma}_{\mathbf{B}}$,

where $c = \left(N(G-1) \right)^{-1} \left(N^2 - \sum_{g}^{G} n_g^2 \right)$ is the averaged group size. In a balance-design case

(i.e. all higher-level units have the same group size), MUML is the same as the original

unbiased ML estimator. But in an unbalance-design case, MUML is the simplified version

of ML estimation method and only uses a common group size, c, as the weighting scalar

of the between-level variance component in the likelihood function, that is,

$$F_{MUML}\left(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}} \right) = F_{MUML}\left(\mathbf{S}, \hat{\mathbf{\Sigma}} \right) = G \left\{ \ln \left| \hat{\mathbf{\Sigma}}_{\mathbf{W}} + c\hat{\mathbf{\Sigma}}_{\mathbf{B}} \right| + tr \left(\left(\hat{\mathbf{\Sigma}}_{\mathbf{W}} + c\hat{\mathbf{\Sigma}}_{\mathbf{B}} \right)^{-1} \mathbf{S}_{\mathbf{B}} \right) - \ln \left| \mathbf{S}_{\mathbf{B}} \right| - p \right\}$$

$$+ \left(N - G \right) \left\{ \ln \left| \hat{\mathbf{\Sigma}}_{\mathbf{W}} \right| + tr \left(\hat{\mathbf{\Sigma}}_{\mathbf{W}}^{-1} \mathbf{S}_{\mathbf{PW}} \right) - \ln \left| \mathbf{S}_{\mathbf{PW}} \right| - p \right\} \qquad (2.45)$$

So, MUML is also called as limited information or quasi-maximum likelihood

estimation because it assumes that all groups have equal group size, even though they do

not. One of the nice features of MUML is that researcher can use the multi-group analysis

routine provided in traditional latent variable statistical software to conduct the multilevel

analysis. Researchers just need to separate the original data into two groups: the

higher-level group with between-level variance-covariance matrix $\mathbf{S}_{\mathbf{B,MUML}}$ and group

size G (i.e. the original group number), and the lower-level group with within-level

variance-covariance matrix $\mathbf{S}_{\mathbf{PW,MUML}}$ with sample size $N - G$. The multilevel data can

then be analyzed with the multi-group routine. The detailed steps of this process was

provided in Heck and Thomas (2008), and Muthén (1991; 1994). Comparing to FIML,

MUML is simpler in computing the parameter estimates. Muthén and Satorra (1995)

concluded that MUML generally performs equally well as FIML in various conditions;

however, a new simulation study showed FIML has more accurate parameter estimates

than MUML does (Hox, Maas & Brinkhuis, 2008) .

   Another feasible estimation for clustered data is the full information maximum

likelihood estimation, FIML/FML (Arbuckle, 1996; Enders & Bandalos, 2001; Yuan &

Hayashi, 2005). We have already known that FIML estimation was introduced to handle

the incomplete dataset in the previous section. Mehta and Neale (2005) gave a clear

description of why and how FIML can handle the missingness in incomplete data. The

individual-specific likelihood function in Equation(2.26) can be interpreted as the

group-specific likelihood function with n_g individuals inside, and each individual has

p measured variables. With this ability, FIML estimators are more robust for unbalanced

designs (du Toit & du Toit, 2008). For multilevel data with unequal group sizes, the

group-specific likelihood function of each group can fully utilize all the information in the

group, and the summation of group-specific likelihood functions provide us full

information when we maximize it to have the parameter estimates. In detail, the

model-implied variance-covariance matrix and/or mean structure can be constructed for

each group. Then, the discrepancy function between model-implied variance-covariance

matrix and raw data in each group will be formulated as shown in Eq(2.26). The direct

log-likelihood function summarizes group-wise log-likelihood functions across the entire

sample as in Eq(2.27). Finally, we maximize Eq(2.28) to have the vector of estimated

unknown parameters $\tilde{\boldsymbol{\theta}}$ to be the consistent and efficient estimates of unknown parameters.

  With the ability of dealing with missing data and unbalanced data, FIML is a flexible estimation in SEM, such as random slopes analysis (e.g. the dimension and the variability of different group-specific variance-covariance matrix can be individually specified) (Heck & Thomas, 2008; Mehta & Foorman, 2005). For longitudinal data analysis, FIML is also good for the unequal distant repeated measures in latent growth curve modeling with the ability in dealing with missing data (Duncan, Duncan, & Hops, 1996; Mehta & Neale, 2005; Wu, West, & Taylor, 2009).

3. USING STRUCTURAL EQUATION MODELING TO ANALYZE COMPLEX SURVEY DATA: A COMPARISON BETWEEN DESIGN-BASED SINGLE-LEVEL AND MODEL-BASED MULTI-LEVEL APPROACHES

Although the design-based approach is relatively simpler for model specification, it presumes that the within-level and between-level models are exactly the same which may not always be true. On the other hand, the advantage of using multilevel model is the flexibility of specifying different models for different levels. Indeed, Muthén and Satorra (1995) showed that these two approaches (design-based vs. model-based) performed equally well on analyzing complex survey data with exactly the same model structure across all data levels.   However, design factors such as the structure/equality of the within- and between-level models and evaluation criteria such as the coverage of the parameter estimates and the empirical power for detecting the parameter estimates were not considered in their simulation.

In this study, we extended Muthen and Satorra's (1995) findings by comparing the two multilevel modeling approaches (i.e., design-based versus model-based approaches) ~~on~~ for analyzing multilevel data with the consideration of a set of design factors including number of clusters, cluster size, intra-class correlation (ICC), and the structure/equality of

the between-level and within-level models. We adopted Mplus (V5.2; Muthén & Muthén,

2007) for all the data generations and analyses. Mplus (V5.2) has the built-in routines (i.e.,

TYPE=COMPLEX and TYPE=TWOLEVEL) for analyzing multilevel data with the two

approaches. The TYPE=COMPLEX routine is used for the design-based approach in

which only one (single-level) model is needed for specification while the

TYPE=TWOLEVEL routine is used for the model-based approach which allows

researchers to specify different models for different levels of data. By default, both routines

use maximum likelihood parameter estimator and robust sandwich standard error estimator

in which the formula for calculating the variance components includes a score factor

"sandwiched" between two copies of the Hessian matrix (Hardin & Hilbe, 2007). This

estimation procedure is termed Maximum Likelihood Estimation with robust standard

error correction (MLR) in Mplus, which is useful for non-normality and non-independence

of observations, and the corresponding chi-square test statistic is asymptotically equivalent

to the Yuan-Bentler T2* test statistic (Muthén & Muthén, 2007). The robust parameter

estimator is also called as Huber-White robust standard error estimate, survey variance

estimate, design-based variance estimate, and empirical variance estimate and has been

widely used in survey statistics (Hardin & Hilbe, 2007). With the use of this robust

estimator, an asymptotically consistent estimate of covariance matrix can be derived free

from distributional assumptions of observations (Huber, 1967; White, 1982; Hardin &

Hilbe, 2007).

There is another feasible but rarely mentioned modeling strategy for complex

survey data named the "maximum model" (Hox, 2002). In the maximum model strategy,

a saturated between-level model is specified; that is, all the unique elements in the

between-level variance-covariance matrix are estimated with the consumption of all

available degrees of freedom in the higher-level. Originally suggested as the baseline

model before any further higher level model construction with theoretical evidence, the

maximum model has been discussed by several researchers (e.g., Hox, 2002; Stapleton,

2006; Yuan & Bentler, 2007). Nevertheless, the performance of this modeling strategy

has not yet been systemically examined. The purpose of this study is to compare the

potential differences of analyzing multilevel data with a design-based single-level

confirmatory factor analytic (CFA) model and the two model-based multilevel CFA

models (i.e., the two-level true model and the maximum model) on the overall model

chi-square test and several commonly used fit indices, the parameter estimates, 95%

coverage for both fixed-effect and random-effect and the respective statistical inferences.

Specifically, our major research question is to investigate what the effects of the number of

clusters, cluster size, intra-class correlation (ICC), and model specification are on the

overall model fit indices, the fixed-effect and random-effect estimates, the 95% coverage

rate, and respective statistical inferences when:

    1)   the between-level and within-level have the same model structure;

    2)   the between-level and within-level have different model structures, including

        a) Complex within-level and simple between-level structure, and

        b) Simple within-level and complex between-level structure.

The three settings are commonly found in empirical research. Caprara, Barbaranelli,

Borgogni, and Steca (2003) tested the relation between efficacy beliefs and teachers' job

satisfaction in an equal between- and within-level SEM model. Example of a complex

between- and simple within-level structure can be found in Beets and Foley (2008), where

the relation between father involvement and neighborhood quality within kindergarteners'

physical activity was examined in a two-factor between-level model with structural

relationship and a one-factor within-level model with covariates predicting the underlying

factor, or in Frenzel, Goetz, Ludtke, and Pekrun (2009), where the relation between teacher

and student enjoyment was investigated in a three-factor within-level structural model and

a four-factor between-level mediation model. A simple between-level and complex

within-level scenario will be a reversed scenario of simple between-level and complex

within-level structure.

3.1 Method

Three simulation scenarios were composed to answer the above research questions,

that is, equal between and within structures for Scenario 1, complex within and simple

between structures for Scenario 2, and simple within and complex between structures for

Scenario 3. In each scenario, a two-level multilevel CFA (MCFA) model were constructed

based on the MCFA model examined by Yuan and Bentler (2002) with some modifications

referring to Hu and Bentler's research (1998, 1999). Four factors were controlled when

conducting simulation in each scenario: cluster number (CN=50, 150 & 300;Hox & Maas,

2001; Muthén & Satorra, 1995), cluster size (CS=10, 50, and 200; Hox & Maas, 2001),

intra-class correlation (low ICC=0.1 & high ICC = 0.5,; Hox & Maas, 2001), and model

specification. The simulation parameters were chosen based on the following criterions: a)

Simulation parameters were used in previous simulation studies, b) simulation parameters

were set based on the real data conditions, such as CN=150 (e.g. Hox and Maas (2001)

mentioned , in multilevel modeling, it can be difficult to obtain data from as many as 200

groups), and c) to maximize the difference between outcomes of the various simulation

conditions and the effect of imbalance, the simulation parameters were chosen to be as

different as possible, such as ICC=0.1 and 0.5 (e.g. Hox & Maas (2001) mentioned, in

educational research, most ICC's are below 0.20. So, ICC=0.1 was the low ICC condition

and ICC =0.5 was a much higher value for the high ICC condition) Monte Carlo procedure

of Mplus 5.2 (Muthén & Muthén, 2009) was used to produce 1000 replications for each

combination of factors in each scenario, that is, a total of: 3 (scenarios) $\times$3 (cluster numbers)

$\times$3 (cluster sizes) $\times$2 (ICCs) $\times$ 1000 = 54,000 replications were generated.

In model specification, one single-level MLR model using TYPE=COMPLEX, and

two multi-level MLR models using TYPE=TWOLEVEL were constructed. The one-level

model had correct specification for the within-level structure without modeling the

between-level structure. The two two-level MLR models had correct specification for the

within structure but one of the two-level models had correctly specified between-level

structure while the other had a saturated between-level structure (i.e. the full rank scaled

between-level covariance matrix was freely estimated). For simplicity to distinguish the

two two-level models and the single-level model, from now on we named the two-level

model with saturated between-level structure as "two-level maximum model," the

two-level model with correct between-level structure as "two-level true model," and the

single-level model with specification only in the within-level structure as "one-level

model". All the generated datasets (i.e., the 54,000 replications) were analyzed by these

three model specifications separately using Mplus 5.2 (Muthén & Muthén, 2007). Detailed

information of each scenario was depicted below.

3.1.1 Scenario 1: Equal between-level model / within-level model

A set of complex survey data were generated based on an equal between- and

within- structure CFA modeling as shown in Figure 2. The within- and between- levels

were specified to have an equal factor structure with nine observed variables and three

common factors. By following previously published simulation studies (e.g., Hox & Maas,

2001; Muthen & Satorra, 1995), the correlations between the common factors were set to

0.3 while most of the pattern coefficients (i.e. a more specific name for factor loadings,

Thompson, 2004) between factors and outcomes were assigned to be 0.8. Two

cross-loaded factor loadings were specified to 0.4 in the within-level model (i.e. FW2→V3

and FW3→V6), so were the cross-loaded factor loadings in the between-level model (i.e.

FB2→Y3, and FB3→ Y6). The residual variances of all manifest variables were taken as

values that would yield unit-variance measured variables under normality (Hu & Bentler,

1998) and were specified to be equal to .36. The two-level maximum model and the

two-level true model were constructed to model the simulated data using the

TYPE=Twolevel routine. On the other hand, a competing one-level model was constructed

with only the specification in the within-level model as shown in Figure 2(ii) to model the

simulated data using the TYPE=Complex routine.
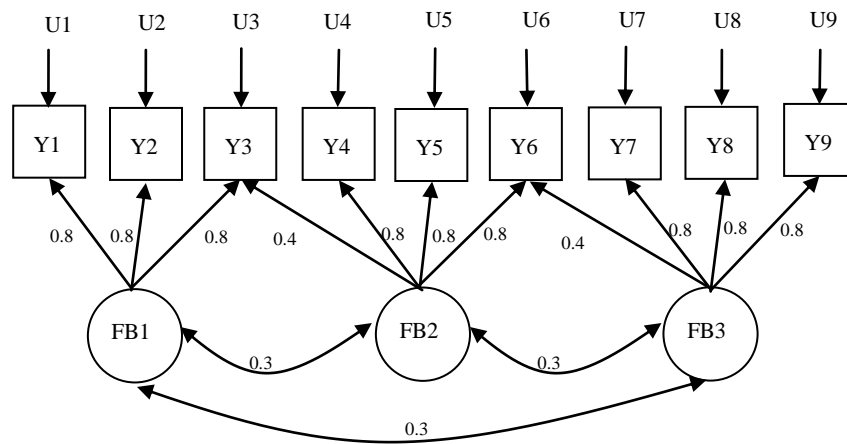
(i)Between-level model
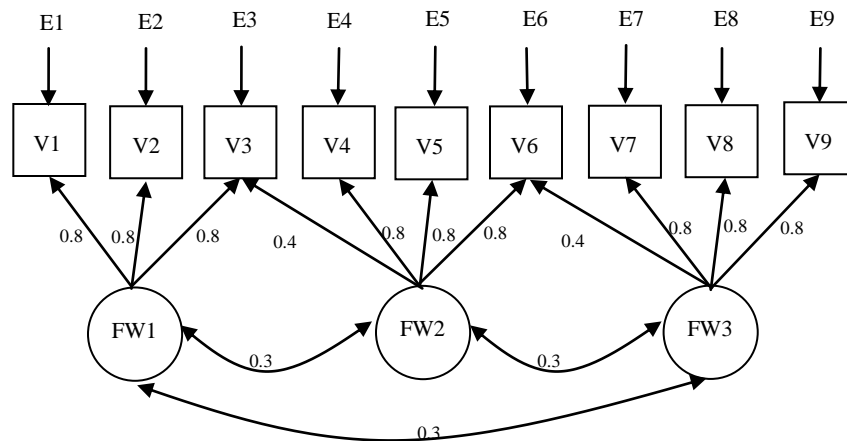
(ii)Within-level model

Figure 2. Simulated multilevel Confirmatory Factor Analysis model for
Scenario 1: Equal Between-level/Within-level Structures.

3.1.2 Scenario 2: Simple between-level model / complex within-level model

The difference of Scenario 1 from Scenario 2 was the complexity of the

between-level model. In Scenario 2 (as shown in Figure 3), the between-level model was

reduced to a one-factor confirmatory measurement model as shown in Figure 3. The

parameterization of the within-level model in Scenario 2 was the same as that of the

within-level model in Scenario 1. In the between-level model, the pattern coefficients

between factor FB and the nine measured variables were fixed at 0.8 and the residual

variances were set as 0.36. Likewise, the two-level true model and the two-level maximum

model were employed to analyze the simulated data using the TYPE=Twolevel routine

while the one-level model was used to model the simulated data using the Type=Complex

routine.
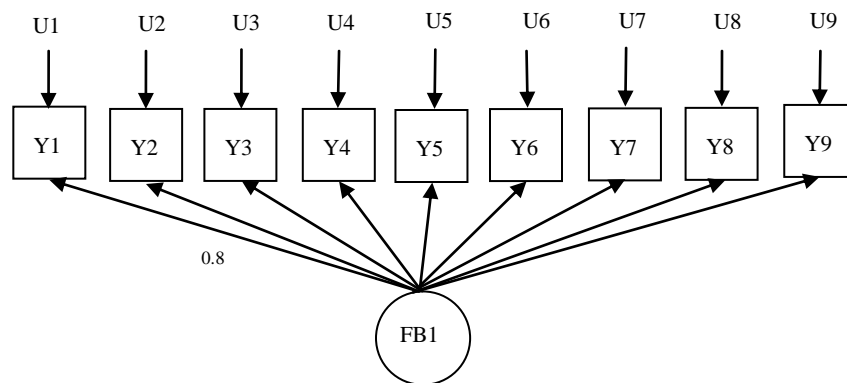
(i)Between-level model
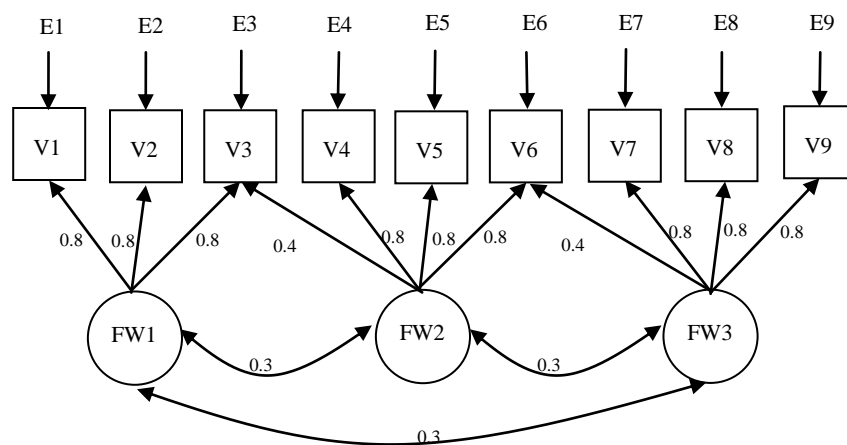
(ii)Within-level model

Figure 3. Simulated multilevel Confirmatory Factor Analysis model for
Scenario 2: Simple Between-level/Complex Within-level Structures.

3.1.3 Scenario 3: Complex between-level model / simple within-level model

Scenario 3 (as shown in Figure 4) was a reversed model of Scenario 2. The

parameterization of the between-level structure in Scenario 3 was the same as that of the

between-level structure in Scenario 1. In the within-level structure, the pattern coefficients

between factor FW1 and the nine measured variables were fixed at 0.8 and the residual

variances were set as 0.36 for the manifest variables. Likewise, the two-level true model

and the two-level maximum model were employed to analyze the simulated data using the

TYPE=Twolevel routine while the one-level model was used to model the simulated data

using Type=Complex. The results from the three models were compared and discussed.
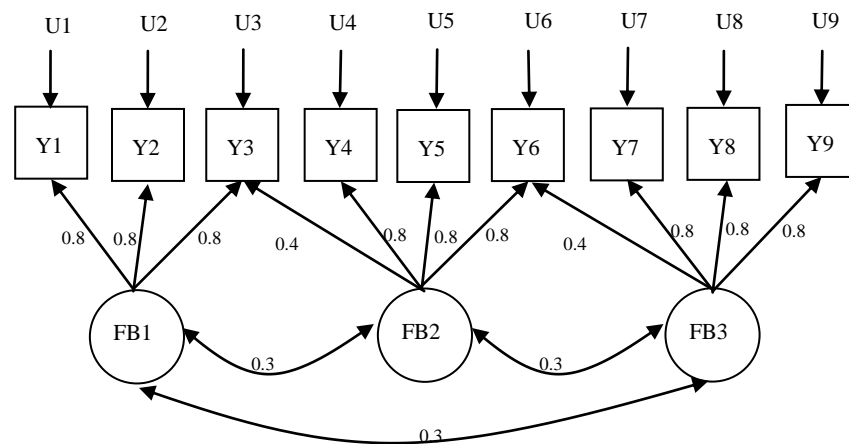
(i) Between-level model
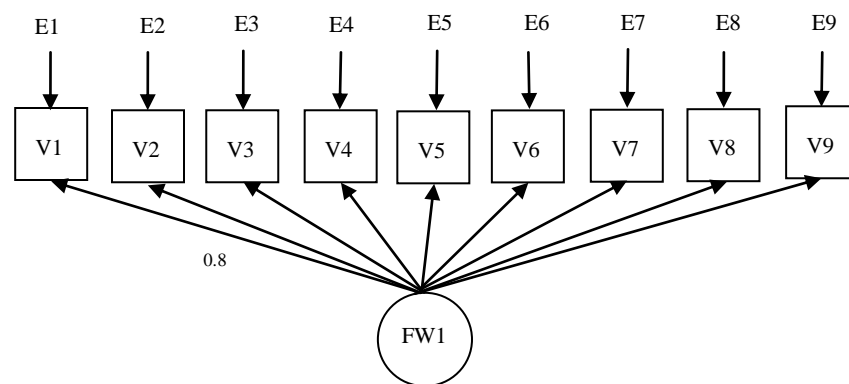
(ii)Within-level model

Figure 4. Simulated multilevel Confirmatory Factor Analysis model for
Scenario 3: Complex Between-level/ Simple Within-level Structures.

3.2 Results

3.2.1 Scenario 1: Equal between-level model / within-level model

The convergence rate of analyses equaled one in all sample size and ICC settings of all the three modeling strategies.

Evaluation of Test Statistic and Model Fit Indices. Values of overall model chi-square test statistics shown as χ^2, CFI, RMSEA, and SRMR for different combination of cluster number and cluster size by ICC were compared in Table 1. In both the low and high ICC settings, the one-level model χ^2 values, which are asymptotically equivalent to the Yuan-Bentler T2* test statistic, were roughly half those for the two-level true model across different combinations and were close to the theoretical value. That is, the χ^2 values reflected the difference in the degrees of freedom; the *df* for the one-level model and the two-level maximum model was 22, which was half the *df* for the two-level true model. When ICC =0.5, the two-level maximum model, which used up all the *df* in the between-level model, produced a more consistent test statistic which was closer to the df of within-level model than the one-level model (e.g. when [CN, CS, ICC]=[50, 10, 0.5],

$\chi^2_{\text{Maximum Model}} = \chi^2_{\text{Maximummodel}} = 23.4$ and $\chi^2_{\text{One-level MLR}} = \chi^2_{\text{one-levelMLR}} = 26.5$).

As for the model fit indices, all modeling methods showed adequate model fit to the dependent data with equal between/within structure. All models had CFI greater than 0.99 and RMSEA smaller than 0.019. Two-level true model and two-level maximum model, in particular, had the same CFI values which were very close to 1, indicating a perfect fit of the model to the dependent data. The CFI values of the one-level model were smaller than the CFI values of the other two-level models as ICC increased, which indicated more lack of fit of the one-level model to the data when between-level structure was not modeled. The one-level model also had higher RMSEA values than the two-level true model and the two-level maximum model across all combinations of cluster number and cluster size, especially at ICC=0.5. The highest RMSEA across all models was 0.019 which occurred at the setting [CN, CS, ICC]=[50, 10, 0.1] of the two-level true model. The RMSEA for the two-level maximum model was the smallest or at least was equal to those of the other two models. SRMR-between and SRMR-within were reported for the two-level true model and two-level maximum model while the one-level model reported a single SRMR value. The SRMR-within for the two-level true model and the two-level maximum model was lower than the SRMR-within of their one-level counterpart.

Table 1. Test Statistic and Model Fit Indices for Scenario 1: Equal
Between-level/Within-level Structures

Model	χ^2	CFI	RMSEA	SRMR	
			ICC=0.1		
50(10)				Between/within	
Two-level	59.980	0.998	0.019	0.086	0.017
One-level	25.409	0.994	0.015	----	0.023
Maximum model	25.443	0.998	0.015	0.006	0.017
50(200)					
Two-level	48.749	1.000	0.003	0.068	0.004
One-level	22.603	0.995	0.004	----	0.017
Maximum model	22.675	1.000	0.002	0.000	0.004
300(10)					
Two-level	45.633	1.000	0.004	0.033	0.007
One-level	22.659	0.999	0.004	----	0.010
Maximum model	22.272	1.000	0.004	0.002	0.007
300(200)					
Two-level	44.418	1.000	0.001	0.027	0.001
One-level	22.082	1.000	0.001	----	0.007
Maximum model	22.216	1.000	0.001	0.000	0.001
			ICC=0.5		
50(10)					
Two-level	50.279	0.998	0.015	0.068	0.020
One-level	26.544	0.992	0.017	----	0.031
Maximum model	23.380	0.998	0.012	0.002	0.020
50(200)					
Two-level	48.574	1.000	0.003	0.061	0.004
One-level	26.807	0.994	0.004	----	0.027
Maximum model	22.693	1.000	0.002	0.000	0.004
300(10)					
Two-level	45.005	1.000	0.004	0.027	0.008
One-level	22.790	0.999	0.005	----	0.013
Maximum model	22.248	1.000	0.004	0.001	0.008
300(200)					
Two-level	44.291	1.000	0.001	0.024	0.002
One-level	22.147	0.999	0.001	----	0.011
Maximum model	22.226	1.000	0.001	0.000	0.002

Note. Two-level= two-level true model (*df* =44); One-level = one-level model (*df*= 22);
Maximum model = two-level maximum model (*df*= 22). In the model column, 50(10)
represents cluster number=50 and cluster size=10; thus, sample size for this setting equals

500. The same notation is used for the rest of settings. χ^2 = overall model Chi-Square test

statistics; CFI = Comparative Fit Index; RMSEA = Root-Mean-Square of Error
Approximation; SRMR = Standardized Root Mean Square Residual.

Fixed effect estimates. The parameter estimates, 95% coverage rate, and

percentage of significant coefficients of the fixed effects for the smallest and largest

sample size settings were provided in Table 2. The results of the smallest and largest

sample size settings were selected for report because they showed the trend of change in

the criterion variables. Complete output of the results is available from the first author

upon request. In general, the one-level model produced consistent and efficient estimates

of factor loadings in Scenario 1. Specifically, as the sample size became larger, that is, a

larger product of cluster number and cluster size, more consistent and efficient fixed

effects estimates were observed. Pattern coefficient estimates were more efficient at small

ICC settings. As for the two-level true model, the parameter estimates in the within-level

model were more consistent and efficient than the parameter estimates in the

between-level model because of the larger sample size in the within-level model.

Moreover, in the high ICC setting, the between-level model produced more efficient fixed

effect estimates than in low ICC setting. When two-level maximum model was used, the

fixed effect estimates were more consistent and efficient, and the standard errors were

smaller than those from the one-level model and were closer to those from the two-level

true model. The column labeled "95% Cover" gave the proportion of replications for which

the 95% confidence interval contains the true population parameter value. As sample size

increased, the 95% confidence interval coverage rate for the two-level maximum model

was identical to those from the two-level true model.

All settings produced statistically significant fixed effect estimates in the

within-level model of the two-level true model, the two-level maximum model, and the

one-level model with the value of 1.000 shown in the column labeled "% Sig Coeff."

However, in the between-level model of the two-level true model, due to the inconsistent

and inefficient fixed effect estimates, the empirical percent of statistically significant

parameter estimates deviated from 100. The worst result occurred at the low ICC and small

sample size settings (e.g. for [CN, CS, ICC]=[50, 10, 0.1] of two-level true model, % of Sig.

Coeff. was equal to 0.549 for Y5, 0.203 for Y6 and 0.126 for Y3).

Table 2. Fixed Effects Estimates of Scenario 1: Equal Between-level/Within-level Structures

Fixed effect	Two-level true model						One-level model			Two-level maximum Model		
	Estimates		95% Cover		% Sig. Coeff.		Estimates	95% Cover	% Sig. Coeff.	Estimates	95% Cover	% Sig. Coeff.
ICC=0.1 50(10)												
FB2/FW2 by Y4/V4	between	within	between	within	between	within						
	.800	0.800					0.800			0.800		
	(.000)	(.000)	1.000	1.000	0.000	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5	0.862	0.799					0.804			0.799		
	(2.07)	(.033)	0.953	0.940	0.549	1.000	(.066)	0.957	1.000	(.033)	0.944	1.000
Y6/V6	0.898	0.800					0.804			0.800		
	(2.24)	(.035)	0.969	0.947	0.203	1.000	(.067)	0.936	1.000	(.034)	0.938	1.000
Y3/V3	0.385	0.401					0.403			0.401		
	(2.17)	(.030)	0.984	0.954	0.126	1.000	(.053)	0.947	1.000	(.030)	0.935	1.000
300(200)												
FB2/FW2 by Y4/V4	between	within	between	within	between	within						
	0.800	0.800					0.800			0.800		
	(.000)	(.000)	1.000	1.000	0.000	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5							0.800			0.800		
	0.809	0.800					(.021)			(.003)		
	(.101)	(.003)	0.955	0.946	1.000	1.000		0.948	1.000		0.946	1.000
Y6/V6							0.800			0.800		
	0.813	0.800					(.020)			(.003)		
	(.150)	(.003)	0.962	0.947	0.999	1.000		0.942	1.000		0.947	1.000
Y3/V3							0.400			0.400		
	0.397	0.400					(.014)			(.003)		
	(.175)	(.003)	0.961	0.946	0.657	1.000		0.952	1.000		0.946	1.000
ICC=0.5 50(10)												
FB2/FW2 by Y4/V4	between	within	between	within	between	within						
	0.800	0.800					0.800			0.800		
	(.000)	(.000)	1.000	1.000	0.000	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5	0.828	0.799					0.807			0.800		
	(.185)	(.048)	0.935	0.941	0.995	1.000	(.082)	0.949	1.000	(.048)	0.944	1.000
Y6/V6	0.825	0.801					0.805			0.802		
	(.208)	(.048)	0.937	0.946	0.994	1.000	(.088)	0.940	1.000	(.051)	0.930	1.000
Y3/V3	0.401	0.400					0.399			0.400		
	(.178)	(.043)	0.941	0.944	0.739	1.000	(.075)	0.934	1.000	(.044)	0.930	1.000
300(200)												
FB2/FW2 by Y4/V4	between	within	between	within	between	within						
	0.800	0.800					0.800			.800		
	(.000)	(.000)	1.000	1.000	0.000	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5	0.802	0.802					0.800			0.800		
	(.060)	(.004)	0.946	0.949	1.000	1.000	(.030)	0.951	1.000	(.004)	0.949	1.000
Y6/V6	0.804	0.799					0.801			0.800		
	(.064)	(.005)	0.939	0.952	1.000	1.000	(.032)	0.937	1.000	(.005)	0.952	1.000
Y3/V3	0.402	0.399					0.401			0.400		
	(.055)	(.004)	0.945	0.942	1.000	1.000	(.027)	0.942	1.000	(.004)	0.942	1.000

Note: Standard error (SE) of parameter estimates were shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contains the true population parameter value; % Sig Coeff = proportion of replications for which produced statistically significant estimates. The pattern coefficients loaded on the between-level factor FB2 of Y4, Y5 and Y6 were fixed at 0.8, and Y3 was set at 0.4. The pattern coefficients loaded on within-level factor FW2 of V4, V5 and V6 were fixed at 0.8, and V3 was set at 0.4 in the within-level model.

Random effect estimates. The residual variance, factor variance, and covariance had the same pattern of simulation result across all sample size settings in the two multilevel models and the one-level model in Scenario 1. In addition, the variance pattern was similar regardless of high or low ICCs. For illustration purpose, Table 3 showed the result of factor covariance when cluster number=300 and cluster size=200 to demonstrate the overall random effect findings. The two-level true model and the two-level maximum model had consistent estimation of factor covariance to the population value. All the factor covariance estimates were around 0.300 at both between and within-level models, and became more consistent and efficient as sample size increased.

However, the factor covariance estimates from the one-level model, where the nested structure was accounted for by using a single-level model assuming equal between and within structures with robust standard error correction, told a different story. The factor covariance estimate was around 0.600, which was twice as large as that in the two-level true model, across different combination of cluster number and cluster size in the one-level model. In other words, the one-level model analysis estimated the total factor covariance component without separating the between and within factor variance portions while two two-level models produced level-specific variance component estimates. When it comes to

the 95% confidence interval coverage rate, the two-level true model produced coverage

rates close to 95%; whereas, the one-level model had poor coverage rate since the one-level

model estimated the total factor covariance, and the confidence interval seldom captured

the true factor covariance, which was divided into the between- and within-level

components respectively. Moreover, the 95% coverage rate differed across different

simulation settings in one-level model. The standard errors of random effect estimates in

the high ICC setting were larger than those in the low ICC setting. Therefore, the 95%

confidence intervals in ICC=0.5 were constructed with larger margin of error, which in

turn erroneously gave a large nonzero coverage rate in the high ICC scenario, especially

when sample size was small (e.g. in [CN, CS, ICC]=[50, 10, 0.5], 95% coverage rate of

$\hat{\Psi}_{FW1,FW3}$=0.702).

For large sample sizes,, the percent of significant coefficients was close to 1.00 for

the three models, indicating the covariance estimates were far away from zero. However,

for the two-level true model in the small sample size setting, the between-level random

effect estimates were inefficient with inflated standard error. The empirical rate of

significant coefficients was underestimated. Even in the high ICC setting, the standard

error of random effect estimate was still large, which result in underestimated percent of

statistically significant coefficients(e.g. when [CN, CS, ICC]=[50, 10 , 0.5], the % of Sig.

Coeff. =0.247 for $\hat{\Psi}_{12}$, 0.268 for $\hat{\Psi}_{13}$ and 0.287 for $\hat{\Psi}_{23}$).

The two-level maximum model produced more consistent random effect estimates

which were close to the within-level population values in the two-level true model.

Furthermore, the two-level maximum model had more efficient random effect estimates

than the one-level model, and the standard errors of random effect estimates were close to

those for the within-level model in the two-level true model especially in the small sample

size and low ICC setting.

Table 3. Comparison of Factor Covariances for Large and Small Sample Size Setting in Scenario 1: Equal Between-level/Within-level Structures

Covariance	Two-level true model						One-level model			Two-level maximum Model		
	Estimates		95% Cover		% Sig. Coeff.		Estimates	95% Cover	% Sig. Coeff.	Estimates	95% Cover	% Sig. Coeff.
ICC=0.1												
50(10)	Between	Within	Between	Within	Between	Within						
F1 with F2	0.304 (.199)	0.305 (.099)	0.947	0.945	0.477	0.867	0.592 (.138)	0.447	0.998	0.303 (.096)	0.931	0.889
F1 with F3	0.297 (.174)	0.306 (.097)	0.945	0.955	0.540	0.905	0.596 (.133)	0.382	1.000	0.305 (.094)	0.950	0.915
F2 with F3	0.298 (.198)	0.303 (.098)	0.942	0.945	0.474	0.886	0.595 (.137)	0.406	0.998	0.301 (.095)	0.935	0.891
300(200)	Between	Within	Between	within	between	within						
F1 with F2	0.302 (.041)	0.300 (.008)	0.926	0.947	1.000	1.000	0.597 (.035)	0.000	1.000	0.300 (.008)	0.947	1.000
F1 with F3	0.303 (.041)	0.300 (.008)	0.934	0.937	1.000	1.000	0.600 (.033)	0.000	1.000	0.300 (.008)	0.937	1.000
F2 with F3	0.299 (.041)	0.302 (.008)	0.930	0.934	1.000	1.000	0.598 (.034)	0.000	1.000	0.302 (.008)	0.934	1.000
ICC =0.5												
50(10)	Between	Within	Between	Within	Between	Within						
F1 with F2	0.287 (.206)	0.301 (.061)	0.936	0.932	0.247	1.000	0.591 (.206)	0.748	.853	0.301 (.060)	0.932	1.000
F1 with F3	0.293 (.202)	0.301 (.059)	0.943	0.956	0.268	1.000	0.597 (.199)	0.702	.894	0.300 (.060)	0.956	1.000
F2 with F3	0.306 (.199)	0.299 (.060)	0.942	0.937	0.287	1.000	0.609 (.204)	0.696	.879	0.299 (.060)	0.937	1.000
300(200)	Between	within	Between	within	between	within						
F1 with F2	0.300 (.076)	0.300 (.005)	0.934	0.947	0.986	1.000	0.600 (.076)	0.014	1.000	0.300 (.005)	0.956	1.000
F1 with F3	0.304 (.074)	0.300 (.005)	0.946	0.949	0.989	1.000	0.604 (.073)	0.010	1.000	0.300 (.005)	0.937	1.000
F2 with F3	0.297 (.075)	0.300 (.005)	0.939	0.955	0.985	1.000	0.598 (.075)	0.016	1.000	0.300 (.005)	0.935	1.000

Note: Standard error of parameter estimates were shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contains the true population parameter value; % Sig Coeff = proportion of replications for which produced statistically significant estimates. The factor covariance was set at 0.3.

3.2.2 Scenario 2: Simple between-level model / complex within-level model

The convergence rate of analyses was equal to 1.0 in all sample size and ICC settings of the one-level model,, the two-level maximum model, and in all sample size and high ICC settings of the two-level true model, while when the cluster number decreased to 50 and ICC equaled 0.1, the convergent rate started to decrease in the two-level true model (e.g. when CN=50, the convergence rate = 100% for CS=200, 99% for CS=50 and 95% for CS=10).

Evaluation of Test Statistic and Model Fit Indices. The model fit indices showed few differences among the one-level model,, the two-level true model, and maximum model. All model fit indices showed adequate model fit. Overall model chi-square test statistic shown as χ^2, CFI, RMSEA, and SRMR for the three model under different simulation settings are compared in Table 4. The χ^2 values had a different pattern from those for the first scenario. Due to the unequal between- and within- level structures, χ^2 values in the one-level model deviated from their degrees of freedom (the theoretical χ^2

value), which was 22 for the one-level model. However, the two-level maximum model

(*df*=22) and two-level true model (*df*=49) still had χ^2 values close to their degree freedom

across all settings. CFIs showed good fit of the one-level model and failed to give a clue to

the different structures in the between-level and within-level model. The same result was

found for RMSEAs. SRMR-between and SRMR-within were reported for the two-level

models and a single SRMR for one-level model. The SRMRs exhibited good fit for the

models. Only the SRMR-between under the setting of small sample size and low ICC

exceeded the rule of thumb (e.g. SRMR-B=0.098 at [CN, CS, ICC]=[50, 10, 0.1]) for the

two-level true model due to little variance in the between-level model with small sample

size .

Table 4. Test Statistic and Model Fit Indices for Scenario 2: Simple
Between-level/Complex Within-level Structures

Model	χ^2	CFI	RMSEA	SRMR	
				ICC=0.1	
50(10)				Between/within	
Two-level	58.716(12.855)	0.998(0.001)	0.017(0.012)	0.098(0.030)	0.010(0.002)
One-level	30.666(10.728)	0.995(0.005)	0.024(0.017)	----	0.015(0.003)
Maximum Model	23.774(7.203)	0.999(0.002)	0.012(0.014)	0.006(0.001)	0.017(0.004)
50(200)					
Two-level	54.109(10.826)	1.000(0.000)	0.003(0.003)	0.043(0.009)	0.002(0.000)
One-level	34.639(10.146)	0.995(0.005)	0.007(0.004)	----	0.011(0.002)
Maximum Model	22.656 (6.869)	1.000(0.000)	0.002(0.003)	0.000(0.000)	0.0024(0.001)
300(10)					
Two-level	51.136(10.788)	1.000(0.000)	0.004(0.004)	0.045(0.010)	0.004(0.001)
One-level	25.156(8.016)	0.999(0.001)	0.006(0.006)	----	0.011(0.002)
Maximum Model	22.243(6.967)	1.000(0.000)	0.004(0.005)	0.002(0.001)	0.007(0.001)
300(200)					
Two-level	49.416(10.379)	1.000(0.000)	0.001(0.001)	0.017(0.003)	0.001(0.000)
One-level	25.483(8.394)	0.999(0.001)	0.001(0.001)	----	0.009(0.001)
Maximum Model	22.216(6.752)	1.000(0.000)	0.001(0.001)	0.000(0.000)	0.001(0.000)
				ICC=0.5	
50(10)					
Two-level	55.549(11.482)	0.996(0.004)	0.014(0.012)	0.050(0.010)	0.020(0.004)
One-level	28.069 (9.599)	0.993(0.008)	0.020(0.016)	----	0.024(0.004)
Maximum Model	23.250(6.864)	0.999(0.002)	0.011(0.013)	0.002(0.001)	0.020(0.004)
50(200)					
Two-level	53.987(10.904)	1.000(0.000)	0.003(0.003)	0.041(0.008)	0.004(0.001)
One-level	29.733(10.640)	0.995(0.006)	0.005(0.004)	----	0.020(0.004)
Maximum Model	22.703(6.784)	1.000(0.000)	0.002(0.003)	0.000(0.000)	0.004(0.001)
300(10)					
Two-level	50.571(10.500)	1.000(0.000)	0.004(0.004)	0.020(0.003)	0.008(0.002)
One-level	33.763 (9.976)	0.998(0.002)	0.012(0.006)	----	0.012(0.002)
Maximum Model	22.248(6.907)	1.000(0.000)	0.004(0.005)	0.001(0.000)	0.008(0.002)
300(200)					
Two-level	49.453(10.160)	1.000(0.000)	0.001(0.001)	0.016(0.002)	0.002(0.000)
One-level	37.626(10.235)	0.998(0.001)	0.003(0.001)	----	0.011(0.001)
Maximum Model	22.226 (6.747)	1.000(0.000)	0.001(0.001)	0.000(0.000)	0.002(0.000)

Note. Two-level= two-level true model (*df* =49); One-level = one-level model (*df*= 22);
Maximum model = two-level maximum model (*df*= 22). In the model column, 50(10)
represents cluster number=50 and cluster size=10; thus, sample size for this setting equals

500. The same notation is used for the rest of settings. χ^2 = overall model Chi-Square test

statistics; CFI = Comparative Fit Indx; RMSEA = Root-Mean-Square of Error
Approximation; SRMR = Standardized Root Mean Square Residual.

Fixed Effect estimates. For conciseness, only the fixed effect estimates of FW2 by

V3, V4, V5, and V6 in the within-level model were reported in Table 5 because they

contained both single-loaded (i.e. V4 and V5) and cross-loaded (i.e. V3 and V6) observed

indicators. The factor loading estimates of FB1 by Y3, Y4, Y5, and Y6 in two-level true

model was also reported for illustration purpose. According to results in Table 5, the

two-level true model produced good estimates of pattern coefficients for both the

single-loaded and cross-loaded variables under the correctly-specified models. The

one-level model in the large sample size setting, however, yielded good single-loaded

loadings but inconsistent cross-loaded fixed effects. For example, regardless of high or low

ICCs, the parameter estimates for the cross-loaded observed variables, Y3 and Y6,

deviated negatively from the population values (λ_{Y3FW2} =0.4 and λ_{Y6FW2} =0.8) to a

noticeably degree (e.g. when [CN, CS, ICC] = [50, 10, 0.5] , $\hat{\lambda}_{Y3FW2}$ =0.290 and $\hat{\lambda}_{Y6FW2}$

=0.679; when [CN, CS, ICC] = [50, 10, 0.1], $\hat{\lambda}_{Y3FW2}$ =0.370 and $\hat{\lambda}_{Y6FW2}$ =0.768).

The standard errors of the fixed effect estimates also exhibited distinct patterns. In the two-level true model, the standard errors of the pattern coefficients were larger in the between-level model than those in the within-level model. The standard errors (SEs) of parameter estimates were the largest in the between-level models for small sample size settings. When the dependent data was analyzed with the one-level model, the SEs of the loading estimates became smaller as ICCs decreased. Although there were larger SEs of fixed effects in the high ICC setting for one-level model, the 95% coverage rate of the cross-loaded fixed effects, Y3 and Y6, was still poor due to seriously attenuated factor loading estimates (e.g. when [CN, CS, ICC] = [300, 200, 0.5], 95% cov. rate = 0.023 for Y6 and 0.013 for Y3 in the one-level model). To the contrary, the two-level maximum model gave consistent and efficient fixed effect estimates for both single-loaded and cross-loaded indicator which was unbiased to its corresponding population values.

Table 5. Fixed Effects Estimates of Scenario 2: Simple Between-level/Complex Within-level Structures

Fixed effect	Two-level true model						One-level model			Two-level maximum Model		
	Estimates		95% Cover		% Sig. Coeff.		Estimates	95% Cover	% Sig. Coeff.	Estimates	95% Cover	% Sig. Coeff.
ICC=0.1												
50(10)												
FB1/FW2 by	Between	Within	Between	Within	Between	Within						
Y4/V4	1.028	0.800					0.800			0.800		
	(1.153)	(.000)	0.937	1.000	0.235	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5	1.074	0.798					0.799			0.799		
	(1.907)	(.034)	0.929	0.940	0.249	0.996	(.020)	0.959	1.000	(.033)	0.944	1.000
Y6/V6	1.008	0.800					0.768			0.801		
	(1.312)	(.036)	0.929	0.935	0.237	0.996	(.023)	0.686	1.000	(.034)	0.938	1.000
Y3/V3	1.091	0.399					0.370			0.401		
	(3.788)	(.030)	0.939	0.956	0.300	0.996	(.019)	0.668	1.000	(.030)	0.935	1.000
300(200)												
FB1/FW2 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.797	0.800					0.800			0.800		
	(.053)	(.000)	0.954	1.000	1.000	0.000	(.000)	1.000	0.000	(.0000)	1.000	0.000
Y5/V5	0.798	0.800					0.799			0.800		
	(.054)	(.001)	0.949	0.948	1.000	1.000	(.018)	0.957	1.000	(.003)	0.945	1.000
Y6/V6	0.799	0.799					0.769			0.800		
	(.054)	(.001)	0.950	0.951	1.000	1.000	(.019)	0.629	1.000	(.003)	0.950	1.000
Y3/V3	0.800	0.400					0.372			0.400		
	(.052)	(.001)	0.945	0.957	1.000	1.000	(.014)	0.500	1.000	(.002)	0.948	1.000
ICC=0.5												
50(10)												
FB1/FW2 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.818	0.800					0.800			0.800		
	(.152)	(.000)	0.948	1.000	1.000	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5	0.819	0.799					0.804			0.800		
	(.153)	(.048)	0.938	0.944	1.000	1.000	(.081)	0.943	1.000	(.047)	0.937	1.000
Y6/V6	0.819	0.802					0.679			0.803		
	(.153)	(.051)	0.937	0.928	1.000	1.000	(.090)	0.680	1.000	(.051)	0.933	1.000
Y3/V3	0.818	0.399					0.290			0.400		
	(.145)	(.044)	0.926	0.930	0.999	1.000	(.079)	0.688	0.948	(.044)	0.928	1.000
300(200)												
FB1/FW2 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.797	0.800					0.800			0.800		
	(.051)	(.000)	0.943	1.000	1.000	0.000	(.000)	1.000	0.000	(.000)	1.000	0.000
Y5/V5	0.798	0.800					0.801			0.800		
	(.051)	(.004)	0.941	0.949	1.000	1.000	(.029)	0.945	1.000	(.004)	0.949	1.000
Y6/V6	0.799	0.799					0.672			0.800		
	(.051)	(.005)	0.939	0.952	1.000	1.000	(.029)	0.023	1.000	(.005)	0.952	1.000
Y3/V3	0.800	0.399					0.289			0.400		
	(.051)	(.004)	0.944	0.942	1.000	1.000	(.025)	0.013	1.000	(.004)	0.942	1.000

Note: The pattern coefficients loaded on the between-level factor FB1 of Y4, Y5, Y6 and Y3were fixed at 0.8. The pattern coefficients loaded on within-level factor FW2 of V4, V5 and V6 were fixed at 0.8, and V3 was set at 0.4 in the within-level model.

Random effect estimates. The result of random effect estimates for Scenario 2 had

similarities as well as differences from Scenario 1. The result of random effects of [CN,

CS]=[200, 300] was shown in Table 6. To start with the similarities, neglecting the higher

level model specification caused the re-distribution of factor variance, that is, one-level

model reported the total factor variance instead of separate between and within variance

components, while the two-level maximum model produced almost identical results to the

within-level random effect estimates in the two-level true model. As for residual variance,

single-loaded indicator residual variances were close to the summation of residual variance

estimates in between- and within-level regardless of ICCs. However, cross-loaded

indicators, in the high ICC setting in particular, had larger estimate of residual variance

which was no longer close to the summation of residual variances' estimates of two-level

indicators (e.g. when ICC =0.5, $\hat{\theta}_{Y3}$=0.795 in the one-level model while $\hat{\theta}_{Y3, \text{bewteen}}$ =0.357

and $\hat{\theta}_{Y3, \text{within}}$ =0.360 in the two-level true model). The larger residual variance of

cross-loaded indicators resulted from the inaccurate estimates of the pattern coefficients

between cross-loaded indicators and their underlying factors as ICC increased. As for

factor covariance, when the one-level model was used to analyze the simulated data true

structure with one common factor between-level structure and three common factors

within-level structure, its factor covariance estimates were not close to the estimates of the

within-level factor covariance in the other two two-level models, either.

Furthermore, with regard to the efficiency of random effect estimation in the

one-level model, the estimates of factor variance with more cross-loaded indictors was

more efficient than other factors (e.g. when ICC=0.5, SE($\hat{\Psi}_{22}$)=0.129, SE($\hat{\Psi}_{11}$) =0.132,

and SE($\hat{\Psi}_{33}$) =0.131). The same pattern occurred when ICC was equal to 0.1. However,

the factor covariance between distant factors was more efficiently estimated than the factor

covariance between the factors with more cross-loaded indicators and the other factors (e.g.

when ICC=0.1, SE($\hat{\Psi}_{13}$)=0.028, SE($\hat{\Psi}_{12}$) =0.030, and SE($\hat{\Psi}_{23}$) =0.030). The same pattern

occurred when ICC was equal to 0.5. In the two-level true model, the estimates of the

residual variance of cross-loaded indicators were less efficient than the estimate of

single-loaded ones. The same situation occurred in the one-level model as well. Due to

biased random effect estimates, the one-level model had zero or very small 95% CI

coverage rate (range 0% to 16%). Percent of significant coefficients was equal to 100% for

all three models.

Table 6 Random Effect Estimates of Large Sample Size Setting in Scenario 2: Simple Between-level/Complex Within-level Structures

Random effects	Two-level true model						One-level model			Two-level maximum Model		
	Estimates		95% Cover		% Sig. Coeff.		Estimates	95% Cover	% Sig. Coeff.	Estimates	95% Cover	% Sig. Coeff.
ICC=0.1												
Factor Variance	Between	within	Between	within	between	within						
FB1/FW1	0.207(.050)	1.799(.014)	0.940	0.948	1.000	1.000	2.005(.070)	0.164	1.000	1.799(.014)	0.948	1.000
FW2		1.800(.014)		0.956		1.000	1.999(.065)	0.129	1.000	1.800(.014)	0.956	1.000
FW3		1.800(.014)		0.941		1.000	1.999(.068)	0.163	1.000	1.800(.014)	0.941	1.000
Factor Covariance	Between	Within	Between	within	between	within						
F1 with F2		0.300(.008)		0.946		1.000	0.481(.030)	0.000	1.000	0.300(.008)	0.947	1.000
F1 with F3		0.299(.008)		0.941		1.000	0.477(.028)	0.000	1.000	0.299(.008)	0.937	1.000
F2 with F3		0.300(.008)		0.970		1.000	0.490(.030)	0.000	1.000	0.300(.008)	0.934	1.000
Residual Variance	Between	Within	Between	within	between	within						
Y3/V3	0.357(.033)	0.360(.003)	0.933	0.955	1.000	1.000	0.744(.048)	0.000	1.000	0.360(.003)	0.955	1.000
Y4/V4	0.359(.033)	0.359(.003)	0.951	0.951	1.000	1.000	0.718(.044)	0.000	1.000	0.359(.003)	0.951	1.000
Y5/V5	0.358(.033)	0.360(.003)	0.944	0.953	1.000	1.000	0.720(.044)	0.000	1.000	0.360(.003)	0.953	1.000
Y6/V6	0.357(.033)	0.360(.003)	0.947	0.953	1.000	1.000	0.743(.046)	0.000	1.000	0.360(.003)	0.953	1.000
ICC = 0.5												
Factor Variance	Between	within	Between	within	between	within						
FB1/FW1	1.009(.122)	0.999(.009)	0.943	0.938	1.000	1.000	2.021(.132)	0.000	1.000	0.999(.009)	0.937	1.000
FW2		1.000(.009)		0.962		1.000	2.010(.129)	0.000	1.000	1.000(.009)	0.962	1.000
FW3		1.000(.009)		0.943		1.000	2.003(.131)	0.000	1.000	1.000(.009)	0.943	1.000
Factor Covariance	Between	Within	Between	within	between	within						
F1 with F2		0.300(.005)		0.956		1.000	1.277(.098)	0.000	1.000	0.300(.005)	0.956	1.000
F1 with F3		0.299(.005)		0.937		1.000	1.268(.097)	0.000	1.000	0.299(.005)	0.937	1.000
F2 with F3		0.300(.005)		0.935		1.000	1.289(.098)	0.000	1.000	0.300(.005)	0.935	1.000
Residual Variance	Between	Within	Between	within	between	within						
Y3/V3	0.357(.033)	0.359(.003)	0.930	0.950	1.000	1.000	0.795(.042)	0.000	1.000	0.359(.003)	0.950	1.000
Y4/V4	0.358(.033)	0.359(.003)	0.948	0.958	1.000	1.000	0.711(.039)	0.000	1.000	0.359(.003)	0.958	1.000
Y5/V5	0.359(.033)	0.360(.003)	0.939	0.950	1.000	1.000	0.712(.040)	0.000	1.000	0.360(.003)	0.950	1.000
Y6/V6	0.358(.033)	0.360(.003)	0.951	0.951	1.000	1.000	0.790(.041)	0.000	1.000	0.360(.003)	0.951	1.000

Note: Factor variance was set as 0.2 in between-level model and as 1.8 in within-level model in low ICC setting and was set as 1 in both levels in high ICC setting. The factor covariance was set at 0.3, and the residual variance was set at 0.36.

3.2.3 Scenario 3: Complex between-level model / simple within-level model

The convergence rate was equal to one in all sample size and ICC settings of the

one-level model and two-level maximum model and in all sample size settings for

ICC=0.5 of the two-level true model. However, when cluster number decreased to 50 and

ICC equaled 0.1, the convergence rate decreased in the two-level true model (e.g. at

ICC=0.1 and CN=50, the convergence rate = 96% for CS=200, 97% for CS=50, and 89%

for CS=10).

Evaluation of Test Statistic and Model Fit Indices. Unlike the previous two

scenarios, the model fit test statistic and indices showed noticeable differences between the

one-level model, the two-level true model, and maximum model. Table 7 Overall model

chi-square test statistic (shown as χ^2), CFI, RMSEA, and SRMR for the three models

under different simulation settings in Table 7. For the overall model chi-square test statistic,

the test statistic value of the one-level model started to deviate from the degrees of freedom

($df = 27$) across ICC settings and the disparity became larger as the sample size increased.

In high ICC settings, the larger test statistic values indicated the stricter penalty due to

neglecting the modeling of potential between-level variations when the design-based

approach or the one-level model was used to analyze dependent data of this kind. On the

other hand, the three model fit indices still indicated a fair model fit in low ICC settings.

However, the three model fit indices showed incongruent patterns of lack of fit in the high

ICC settings. CFIs consistently showed lack of fit of the one-level model across all sample

size combinations in high ICC settings and gave a clue to unequal structures in the between

and within model (e.g. CFI=0.76, 0.74, 0.78, and 0.77 respectively across different

combinations of sample size at ICC=0.5 for the one-level). RMSEA gave some

information of discrepancy in the between- and within- level models in the smaller cluster

size settings but not in larger cluster size settings (e.g. RMSEA=0.10 at [CN, CS,

ICC]=[300, 10, 0.5] and 0.11 at [CN, CS, ICC]=[50, 10, 0.5], but RMSEA= 0.03 at both

[CN, CS, ICC]=[300, 200, 0.5] and [CN, CS, ICC]=[50, 200, 0.5]). All SRMRs exhibited

good fit of the model to the data for all three models and failed to provide information

regarding the unequal between/within structure. Only the SRMR-between slightly

exceeded the rule of thumb under the setting of small sample size and low ICC (e.g.

SRMR-B=0.09 at [CN, CS, ICC]=[50, 10, 0.1]) for the two-level true model due to little

variance and small cluster size in the between-level model.

Table 7. Test Statistic and Model Fit Indices for Scenario 3: Complex Between-level/ Simple Within-level Structures.

Model	χ^2	CFI	RMSEA		SRMR
			ICC=0.1		
50(10)				Between/within	
Two-level	55.83	1.00	0.02	0.09	0.01
One-level	44.47	0.99	0.03	----	0.02
Maximum Model	28.62	1.00	0.01	0.00	0.01
50(200)					
Two-level	53.98	1.00	0.00	0.07	0.00
One-level	49.74	0.99	0.01	----	0.02
Maximum Model	27.93	1.00	0.00	0.00	0.00
300(10)					
Two-level	49.86	1.00	0.00	0.03	0.00
One-level	95.88	0.99	0.03	----	0.01
Maximum Model	27.02	1.00	0.00	0.00	0.00
300(200)					
Two-level	50.01	1.00	0.00	0.03	0.00
One-level	120.65	0.99	0.01	----	0.01
Maximum Model	27.28	1.00	0.00	0.00	0.00
			ICC=0.5		
50(10)					
Two-level	55.68	1.00	0.01	0.07	0.01
One-level	187.07	0.76	0.11	----	0.08
Maximum Model	28.61	1.00	0.01	0.00	0.01
50(200)					
Two-level	54.21	1.00	0.00	0.06	0.00
One-level	238.21	0.74	0.03	----	0.08
Maximum Model	27.90	1.00	0.00	0.00	0.00
300(10)					
Two-level	49.69	1.00	0.00	0.03	0.01
One-level	816.44	0.78	0.10	----	0.07
Maximum Model	27.03	1.00	0.00	0.00	0.01
300(200)					
Two-level	49.91	1.00	0.00	0.02	0.00
One-level	1021.83	0.77	0.03	----	0.07
Maximum Model	27.27	1.00	0.00	0.00	0.00

Note. Two-level= two-level true model (*df* =49); One-level = one-level model (*df*= 27); Maximum model = two-level maximum model (*df*= 27). In the model column, 50(10) represents cluster number=50 and cluster size=10; thus, sample size for this setting equals 500. The same notation is used for the rest of settings.

Fixed effect estimates. Fixed effect estimate of pattern coefficients between indicators and the between-level factor (FB2) and within-level factor (FW1) were tabulated in Table 8. Similarly, Y3, Y4, Y5, and Y6 (the between-level manifest variables) and V3, V4, V5, and V6 (the within-level manifest variables) were reported to examine single-loaded as well as cross-loaded indicators. In the small sample size and low ICC settings, the two-level true model generated inconsistent and inefficient estimates of pattern coefficients for both the single-loaded and cross-loaded indicators, except Y4, the marker variable (e.g. $\hat{\lambda}_{Y5FB2}$=1.085, $\hat{\lambda}_{Y6FB2}$=1.015, $\hat{\lambda}_{Y3FB2}$=0.372 at [CN, CS, ICC]=[50, 10, 0.1]). In addition, due to neglecting the modeling of the between-level structure, the one-level model generated biased fixed effect estimates for cross-loaded indicators (Y3 and Y6) which were positively deviated from the population values with the relative bias ranging from 6.3% to 29% as ICC increased from 0.1 to 0.5. Moreover, the pattern

coefficient estimates of single-loaded indicators (Y4 and Y5) were also overestimated and

positively away from the population values with the relative bias ranging from -0.3% to 8%

as ICC increased. The standard errors of the fixed effect estimates also exhibited inflated

patterns as ICC increased in the one-level model. With the larger standard error estimates

in the high ICC setting, the 95% coverage rate of the cross-loaded fixed effects (Y3 and Y6)

was still small due to positively biased point estimates (e.g. when [CN, CS, ICC] = [300,

200, 0.5], 95% CI coverage rate of Y6 = 0.000 and Y3 = 0.022 for one-level model). As the

promising modeling strategy shown in Scenario 2, the two-level maximum model

generated unbiased estimate of the fixed effect and standard error for both single-loaded

and cross-loaded indicators in the within-level model.

Table 8. Fixed Effects Estimates of Scenario 3: Complex Between-level/ Simple Within-level Structures

Fixed effect	Two-level true model						One-level model			Two-level maximum Model		
	Estimates		95% Cover		% Sig. Coeff.		Estimates	95% Cover	% Sig. Coeff.	Estimates	95% Cover	% Sig. Coeff.
ICC=0.1												
50(10)												
FB2/FW1 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.800	0.800					0.799			0.801		
	(0.000)	(0.032)	1.000	0.934	0.000	0.999	(0.048)	0.942	1.000	(0.030)	0.934	1.000
Y5/V5	1.085	0.801					0.800			0.801		
	(1.635)	(0.031)	0.935	0.935	0.540	0.999	(0.048)	0.934	1.000	(0.030)	0.935	1.000
Y6/V6	1.015	0.801					0.853			0.801		
	(1.580)	(0.031)	0.908	0.950	0.269	0.999	(0.050)	0.831	1.000	(0.030)	0.950	1.000
Y3/V3	0.372	0.801					0.855			0.801		
	(1.466)	(0.032)	0.974	0.942	0.140	0.999	(0.053)	0.831	1.000	(0.030)	0.942	1.000
300(200)												
FB2/FW1 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.800	0.800					0.798			0.800		
	(0.000)	(0.003)	1.000	0.949	0.000	1.000	(0.014)	0.936	1.000	(0.003)	0.949	1.000
Y5/V5	0.806	0.800					0.797			0.800		
	(0.100)	(0.003)	0.941	0.955	1.000	1.000	(0.014)	0.945	1.000	(0.003)	0.955	1.000
Y6/V6	0.812	0.800					0.850			0.800		
	(0.146)	(0.003)	0.937	0.938	1.000	1.000	(0.014)	0.051	1.000	(0.003)	0.938	1.000
Y3/V3	0.394	0.800					0.853			0.800		
	(0.169)	(0.003)	0.956	0.942	0.687	1.000	(0.016)	0.066	1.000	(0.003)	0.942	1.000
ICC=0.5												
50(10)												
FB2/FW1 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.800	0.801					0.864			0.801		
	(0.000)	(0.041)	1.000	0.941	0.000	1.000	(0.143)	0.955	1.000	(0.041)	0.939	1.000
Y5/V5	0.825	0.801					0.865			0.801		
	(0.182)	(0.041)	0.947	0.939	0.989	1.000	(0.144)	0.954	1.000	(0.041)	0.939	1.000
Y6/V6	0.841	0.802					1.032			0.802		
	(0.210)	(0.041)	0.958	0.943	0.976	1.000	(0.167)	0.812	1.000	(0.041)	0.943	1.000
Y3/V3	0.413	0.801					0.977			0.801		
	(0.177)	(0.041)	0.959	0.945	0.718	1.000	(0.101)	0.606	1.000	(0.041)	0.943	1.000
300(200)												
FB2/FW1 by	Between	Within	Between	Within	Between	Within						
Y4/V4	0.800	0.800					0.849			0.800		
	(0.000)	(0.004)	1.000	0.952	0.000	1.000	(0.052)	0.893	1.000	(0.004)	0.952	1.000
Y5/V5	0.804	0.800					0.849			0.800		
	(0.060)	(0.004)	0.945	0.949	1.000	1.000	(0.052)	0.884	1.000	(0.004)	0.949	1.000
Y6/V6	0.804	0.800					1.012			0.800		
	(0.064)	(0.004)	0.947	0.935	1.000	1.000	(0.059)	0.022	1.000	(0.004)	0.935	1.000
Y3/V3	0.399	0.800					0.968			0.800		
	(0.055)	(0.004)	0.945	0.938	1.000	1.000	(0.035)	0.000	1.000	(0.004)	0.938	1.000

Note: The pattern coefficients loaded on the between-level factor FB2 of Y4, Y5 and Y6 were fixed at 0.8, and Y3 was set at 0.4. The pattern coefficients loaded on within-level factor FW1 of V4, V5, V6 and V3 were fixed at 0.8 in the within-level model.

Random effect estimates. Likewise, the result of random effect estimates of cluster

size=300 and cluster number=200 was reported for Scenario 3 in Table 9. The

redistribution of random effect was shown similar as the result in Scenarios 1 and 2. That is,

because of neglecting the higher level modeling, the one-level model reported the total

variance component instead of separate between and within variance components, while

the two-level maximum model produced almost identical results, including the variance

component estimates with standard errors and the corresponding statistical inferences as

the within-level random effect estimates in the two-level true model. However, different

from the result of one-level model in Scenarios 1 and 2, the total variance component

estimate of latent factors and the cross-loaded and single-loaded indicators had biased

parameter estimates, which negatively deviated from the summation of variance

components estimates of the between- and within-level models for both ICC settings.

When ICC increased from 0.1 to 0.5, the factor variance estimates of the one-level model

were less efficient and more negatively biased from the population value of total variance

components of the two levels. The relative bias was 4% for ICC=0.1 and -31% for ICC=0.5

(e.g. when [CN, CS, ICC]=[300, 200, 0.5], the $\hat{\Psi}_{11}$ =1.379 with SE = 0.120; $\hat{\Psi}_{11}$ =2.082

with SE =0.058 in [CN, CS, ICC]=[300, 200, 0.1]). The reversed bias pattern occurred for

the residual variance of indicators. When ICC increased from 0.1 to 0.5, the residual

variance estimates of both cross-loaded and single-loaded indicators of the one-level

model were less efficient but more positively biased from the population value of total

variance components with the relative bias ranging from -15% to 32% (e.g. when [CN, CS,

ICC]=[300, 200, 0.5], the $\hat{\theta}_{Y3,\ \text{one-level model}}$ =1.058 with SE= 0.069, which was more biased

and less efficient than the $\hat{\theta}_{Y3,\ \text{one-level model}}$ =0.705 with SE = 0.031 in [CN, CS, ICC]=[300,

200, 0.1]). All the 95% coverage rate of variance component estimates was close to zero

for the biased point estimates.

Table 9. Random Effect Estimates of Large Sample Size Setting in Scenario 3: Complex Between-level/Simple Within-level Structures

Random effects	Two-level true model						One-level model			Two-level maximum Model		
	Estimates		95% Cover		% Sig. Coeff.		Estimates	95% Cover	% Sig. Coeff.	Estimates	95% Cover	% Sig. Coeff.
ICC=0.1												
Factor Variance	Between	within	Between	within	between	within						
FB1/FW1	0.200(0.048)	1.800(0.013)	0.945	0.943	0.997	1.000	2.082(0.058)	0.002	1.000	1.800(0.013)	0.943	1.000
FB2	0.200(0.047)		0.939		0.999							
FB3	0.200(0.043)		0.943		1.000							
Residual Variance	Between	within	Between	within	between	within						
Y3/V3	0.357(0.036)	0.360(0.002)	0.939	0.950	1.000	1.000	0.705(0.031)	0.000	1.000	0.360(0.002)	0.950	1.000
Y4/V4	0.359(0.032)	0.360(0.002)	0.940	0.953	1.000	1.000	0.675(0.029)	0.000	1.000	0.360(0.002)	0.953	1.000
Y5/V5	0.359(0.032)	0.360(0.002)	0.937	0.952	1.000	1.000	0.674(0.029)	0.000	1.000	0.360(0.002)	0.952	1.000
Y6/V6	0.359(0.034)	0.360(0.002)	0.955	0.938	1.000	1.000	0.715(0.032)	0.000	1.000	0.360(0.002)	0.938	1.000
ICC = 0.5												
Factor Variance	Between	within	Between	within	between	within						
FB1/FW1	1.000(0.131)	1.000(0.009)	0.945	0.946	1.000	1.000	1.379(0.120)	0.086	1.000	1.000(0.009)	0.945	1.000
FB2	0.999(0.129)		0.935		1.000							
FB3	0.992(0.128)		0.949		1.000							
Residual Variance	Between	Within	Between	within	between	within						
Y3/V3	0.354(0.048)	0.360(0.002)	0.933	0.947	1.000	1.000	1.058(0.069)	0.000	1.000	0.360(0.002)	0.947	0.000
Y4/V4	0.357(0.044)	0.360(0.002)	0.940	0.954	1.000	1.000	1.007(0.063)	0.000	1.000	0.360(0.002)	0.954	0.000
Y5/V5	0.357(0.044)	0.360(0.002)	0.951	0.951	1.000	1.000	1.009(0.063)	0.000	1.000	0.360(0.002)	0.951	0.000
Y6/V6	0.355(0.046)	0.360(0.002)	0.941	0.937	1.000	1.000	0.939(0.056)	0.000	1.000	0.360(0.002)	0.937	0.000

Note: Standard error of parameter estimates were shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contains the true population parameter value; % Sig Coeff = proportion of replications for which produced statistically significant estimates. Factor variance was set as 0.2 in between-level model and as 1.8 in within-level model in low ICC setting and was set as 1 in both levels in high ICC setting. Residual variance was set at 0.36.

3.3 Discussion

One of the interesting findings from the simulation study is that the overall model

chi-square test and other commonly used fit indices did not provide much helpful

information consistently on the necessity of specifying a different higher level model,

especially when the design-based single-level model was used for the analyzing data with

truly unequal between-level and within-level model structures . In the simple

between-/complex within-level structure scenario (i.e. Scenario 2), although the model fit

test statistic values for the one-level model deviated from the expected values (df=22)

slightly due to unequal between and within structures, the p-value of the overall model

chi-square value was still larger than .05, which led to an incorrect conclusion for the

equivalence of the between-level and within-level models. Similarly, all three fit indices,

namely, RMSEA, SRMR, and CFI were not sensitive to the deviations of the within-level

model from the between-level model based on the traditionally used cutoff criteria (i.e.,

RMSEA <.08, SRMR <.08, and CFI >.95 which indicate an adequately fitted model).

On the other hand, in the third Scenario (i.e., complex-between/simple-within structure), the overall model chi-square test indicated ill fit of the design-based one-level model to multilevel data with complex between-level structure and simple within-level structure. As for the model fit indices, only CFI showed poor model fit of the one-level model to the dependent data under the high ICC condition (i.e., ICC=0.5). Both RMSEA and SRMR in general failed to indicate the lack of fit of the one-level model to the multilevel data with different model structures at different levels. To sum up, overall model chi-square test statistic and CFI model fit index can only offer partial information regarding model misspecification when the design-based approach is used for analyzing the clustered data with simple within- and complex between-level structures.

For the fixed effects estimates, simulation results differed for equal structure and unequal structure scenarios. In the equal-structure scenario, all factor loading estimates were close to the population values in the two-level true model, whether they were uniquely loaded on one factor or crossly loaded on two factors. The coverage rate of 95% confidence interval and the statistical inferences of the estimates for single-loaded and cross-loaded factor loadings in the one-level model (TYPE=COMPLEX routine) were

close to those of the two-level true model and the two-level maximum model (TYPE=TWOLEVEL routine).

However, in unequal-structure scenario, different patterns of result occurred for single-loaded and cross-loaded factor loadings. In the simple between-/ complex within-level scenario, the estimates of single-loaded factor loadings were unbiased estimates to the population values of the within-level model, and the statistical inferences of single-loaded factor loading estimates in the one-level model were close to those of the within-level model in the two-level true model and to those in the two-level maximum model. However, when the one-level model was used for the non-independent data, the estimates of the crossed-loaded indicators were biased. The biased cross-loaded factor loading estimates were exacerbated with increase in ICC. Also, as ICC increased, the standard errors of the fixed effect estimates increased. Nevertheless, the parameter estimates of the cross-loaded observed variables still had relatively low 95% confidence interval coverage rate due to seriously biased pattern coefficient estimates when ICC was high. The situation became worse when the between-level structure was more complex than the within-level structure (i.e. the complex between-/ simple within-level scenario). The pattern coefficient estimates of both single-loaded and cross-loaded indicators

seriously deviated from the population values, and the biases became severer when ICC increased. Meanwhile, the resulted 95% confidence interval coverage rate was also close to zero for biased factor loading estimates when ICC increased.

With regards to random effects, in general, the one-level model only estimated the overall variance components in the equal structure scenario. Moreover, the one-level model accounted for the overall factor covariance as well as the residual variance under the equal structure scenario. For the unequal structures with simple between model and complex within model, the design based one-level model could still produce good estimates of the summation of factor variances in the between- and within-level models. However, the estimate of factor covariance was no longer the combination of factor covariance from each level. Moreover, we observed poorer estimates of cross-loaded indicator residual variance. On the other hand, in the complex between-/simple within-level scenario, the one-level model generated more biased estimate of total factor variance and total indicators residual variances as ICC increased.

To understand the change of random effects, we needed to take the variation of the fixed effect estimates into account. In the equal structure scenario, the fixed effect estimates were asymptotically close to the population values which were set to be the same in the between- and within-level models. With the fixed effect estimates being correctly specified, variance components for the factor variance, factor covariance, and residual variance were re-distributed for the one-level model in which a single variance component was obtained as the sum of the between and within variance components. The estimates for random effects under the equal structure scenario were equivalent to the combination of respective variance components in the between- and within-level models.

However, in the simple-between/complex-within scenario, the situation became more complicated. The quality of the pattern coefficients of the fixed effects differed; the single-loaded pattern coefficient estimates were consistent across all ICC conditions, but the cross-loaded estimates deteriorated especially as ICC increased. Because of the use of the marker variable strategy for model identification, the factor variances were defined with the same metric of the marker variable (e.g., V4 loaded on FW2 in Figure 2 and Table 5), while factor covariance and residual variance were allowed to vary. In order to maintain the consistent amount of indicator variance and to compensate for the inaccurate

cross-loaded factor loadings, the random effect estimates needed to be adjusted as a result.,

Consequently, the factor covariance estimates were no longer close to the summed

between and within factor covariances. Moreover, the unexplained portion of indicator

variance was credited to the residual variance. That is why the cross-loaded indicators had

higher residual variances than single-loaded indicators as shown in the bottom of Table 6.

The situation deteriorated in the complex between-/simple within-level scenario. The

one-level model produced inconsistent factor variance estimates due to the biased

single-loaded pattern coefficient estimate. Especially as ICC increased, the total factor

variance was underestimated because of the inflated estimate of the single-loaded pattern

coefficient (e.g. when [CN, CS, ICC]=[300, 200, 0.5], $\hat{\Psi}_{11}$=1.379, which was negatively

biased from the total factor variance set as 2).

To the contrary, the use of the two-level maximum model recovered the distorted

cross-loaded fixed effect estimates to the population values in the simple

between-/complex within-level scenario, and restored both the distorted single- and

cross-loaded fixed effect estimates to the population values in the complex

between-/simple within-level scenario as well. Thus, we obtained consistent random effect

estimates close to the true variance components in the within-level of the true model.

Another advantage of the two-level maximum model compared to the two-level true model

is that it offers greater power for the lower-level estimates with dependent data of small

sample size. Although the two-level true model specified the multilevel structure of the

data correctly and yielded asymptotically identical lower-level estimate to the generated

population values under most of the conditions, the two-level true model may still produce

inconsistent parameter estimates with inflated standard error, erroneous statistical

inference, and inflated type II error rate, especially under conditions with low ICC and

small sample size.

4. THE EFFECT OF IGNORING DEPENDENCY IN COMPLEX SURVEY DATA FOR CONDITIONED MULTILEVEL LATENT GROWTH CURVE MODELING

Multilevel latent growth curve model (MLGCM), on the other hand, extends the concept of LGCM to include cluster-specific higher-level data in the model. The potential impacts of ignoring the cluster effect/dependency in the longitudinal data by using LGCM rather than MLGCM have not yet been thoroughly investigated because LGCMs used more information in the observed variables than traditional methods (Hancock & Lawrence, 2006) such as its focus on variance, covariance, and mean structure across the time span (Rogosa & Willett, 1985). The major goals of this study were twofold: (1) to examine the effect of considering (model-based MLGCM and design-based MLGCM approaches) or ignoring (regular LGCM approach) the highest data-level on model fit test statistic value, fit indices, between/within regression weights, and between-/within-level factor covariances, residual variances and mean structures, and (2) to compare the effect of incorporating or not incorporating the higher-level covariate in the regular LGCM and design-based MLGCM on the parameter estimates and the statistical inferences. The concepts of LGCM and MLGCM were briefly reviewed, followed by the design and

analysis for the Monte Carlo study. Finally, we discussed implication of the findings with

suggestions for applications.

4.1 Latent Growth Curve Model (LGCM)

Repeated measurements collected from a group of participants can be analyzed

using a LGCM. In the LGCM, all the participants can have their own initial status

(intercept) and growth curve (slope) on the outcome measure. The variability in the initial

status and growth trajectories are modeled as latent variables. The LGCM helps answer

questions about change over time by treating the latent variables as independent, dependent,

control or mediating variables under SEM methodology (Llabre, Spitzer, Siegel, Saab, &

Schneiderman, 2004). Therefore, under the SEM framework, LGCM are considered as a

single-level model in that the random effects are modeled as latent variables (Bovaird,

2007; Curran, 2003).

4.2 Multilevel Latent Growth Curve Model (MLGCM)

We discuss two MLGCM approaches in this paper, design-based MLGCM and

model-based MLGCM. The design-based approach is the single-level LGCM with robust

sandwich standard error estimator (J. W Hardin & J. Hilbe, 2007; Huber, 1967; White,

1980) which takes the sampling scheme of dependent data into consideration when

computing the standard error of parameter estimates (Muthén & Satorra, 1995; Stapleton, 2006). By considering the complex sampling scheme, the empirical standard error estimates can produce consistent statistical inferences. Unlike the design-based MLGCM which adjusts the standard error with the robust standard error estimator given the sampling scheme, the model-based MLGCM builds the multilevel model with level-specific parameter estimates based on the actual levels of sampling units in the complex sampling design (Heck & Thomas, 2008; Luke, 2007). Besides describing the form and pattern of change in a dependent variable over time, Model-based MLGCM can explore the interindividual and intraindividual predictors of this change in which individuals are grouped into clusters that may have structures (Heck & Thomas, 2008; Luke, 2007). An example of two-level repeated measurement data would be the growth of the individual students within the context of classrooms, where repeated measures within a student and the variation in growth parameters among students within a classroom are captured in the first level, and the variation among classrooms is presented in the second level. For example, the experience of teachers (i.e. a classroom-level predictor) may affect the behavioral engagement and academic achievement of students. Ignoring the dependency issue that students are nested in classrooms may result in biased estimation of

the standard error of the fixed effect estimates (or the regression coefficients), which in turn affects the statistical inference of the parameter estimates (Hox, 2002).

Considering a multilevel dataset with T elements of repeated measures (level 1 unit) y_{tig}, $t = 1, 2, ..., T$, of each of I participants (level 2 unit) nested within G groups (level 3 unit). For the ith participant (level 2 unit) within g^{th} group, assuming $\mathbf{y}_{ig}$ is a multivariate normally distributed random vector with T elements of repeated measures (level 1 unit) y_{tig}, $t = 1, 2, ..., T$, that is,

$$
\mathbf{y}_{ig} = \begin{bmatrix} y_{1ig} \\ y_{2ig} \\ \vdots \\ y_{Tig} \end{bmatrix}
\tag{2.46}
$$

where $i = 1, 2, ..., I$ and $g = 1, 2, ..., G$.

Thus, for each gth group (level 3 unit), the random matrix of observations can be arranged as:

$$
\mathbf{y}_g = \begin{bmatrix} \mathbf{y}_{1g} \\ \mathbf{y}_{2g} \\ \vdots \\ \mathbf{y}_{Ig} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} y_{11g} & y_{21g} & \cdots & y_{T1g} \end{bmatrix}' \\ \begin{bmatrix} y_{12g} & y_{22g} & \cdots & y_{T2g} \end{bmatrix}' \\ \vdots \quad \vdots \quad \cdots \quad \vdots \\ \begin{bmatrix} y_{1Ig} & y_{2Ig} & \cdots & y_{TIg} \end{bmatrix}' \end{bmatrix}
\tag{2.47}
$$

Beside on Equation(2.38), the random vector $\mathbf{y}_{ig}$ can be decomposed into its

between-group and within-group components,

$$
\begin{aligned}
\mathbf{y}_{ig} &= \mathbf{y}_{B..g} + \mathbf{y}_{W.ig} \\
&= \boldsymbol{\mu}_B + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B..g} + \boldsymbol{\varepsilon}_{B..g} + \boldsymbol{\mu}_W + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{W.ig} + \boldsymbol{\varepsilon}_{W.ig}
\end{aligned}
\tag{2.48}
$$

Here, the random vector of latent growth factors $\boldsymbol{\eta}_{B..g}$ and $\boldsymbol{\eta}_{W.ig}$ include the latent

growth factors I (intercept factor), S (linear growth factor), S^2 (quadratic growth factor)

and so on to the highest order growth factor S_B^H, and the corresponding factor loadings

matrix $\boldsymbol{\Lambda}_B$ and $\boldsymbol{\Lambda}_W$ are set as follows

$$
\boldsymbol{\eta}_{B..g} = \begin{bmatrix} I_B \\ S_B \\ S_B^2 \\ \vdots \\ S_B^H \end{bmatrix}_{H \times 1}
\quad \text{with } \boldsymbol{\Lambda}_B = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 4 & \dots 2^H \\ 1 & 3 & 9 & \dots 3^H \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{T \times H}
$$

$$
\boldsymbol{\eta}_{W.ig} = \begin{bmatrix} I_W \\ S_W \\ S_W^2 \\ \vdots \\ S_W^H \end{bmatrix}_{H \times 1}
\quad \text{with } \boldsymbol{\Lambda}_W = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 4 & \dots 2^H \\ 1 & 3 & 9 & \dots 3^H \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{T \times H}
$$

The MLGC model can also include different level covariates in the model to

investigate the relationship between latent growth factors and covariates. This is so called

conditional MLGC model. Suppose X (the within-level covariate) and Y (the

between-level covariate) are used to examine the constructs of corresponding growth

factors. Then, the $\boldsymbol{\eta}_{B..g}$ and $\boldsymbol{\eta}_{W.ig}$ in Equation (2.48) can be further expressed as follows:

$$
\boldsymbol{\eta}_{B..g} = \begin{bmatrix} I_B \\ S_B \\ S_B^2 \\ \vdots \\ S_B^H \end{bmatrix} = \boldsymbol{\beta}_{BY} Y + \mathbf{e}_B = \begin{bmatrix} \beta_{0I_B} & \beta_{1I_B} \\ \beta_{0S_B} & \beta_{1S_B} \\ \beta_{0S_B^2} & \beta_{1S_B^2} \\ \vdots & \vdots \\ \beta_{0S_B^H} & \beta_{1I_B} \end{bmatrix} \begin{bmatrix} 1 \\ Y \end{bmatrix} + \begin{bmatrix} e_{I_B} \\ e_{S_B} \\ e_{S_B^2} \\ \vdots \\ e_{S_B^H} \end{bmatrix} \tag{2.49}
$$

$$
\boldsymbol{\eta}_{W..g} = \begin{bmatrix} I_W \\ S_W \\ S_W^2 \\ \vdots \\ S_W^H \end{bmatrix} = \boldsymbol{\beta}_{WX} X + \mathbf{e}_W = \begin{bmatrix} \beta_{0I_W} & \beta_{1I_W} \\ \beta_{0S_W} & \beta_{1S_W} \\ \beta_{0S_W^2} & \beta_{1S_W^2} \\ \vdots & \vdots \\ \beta_{0S_W^H} & \beta_{1I_W} \end{bmatrix} \begin{bmatrix} 1 \\ X \end{bmatrix} + \begin{bmatrix} e_{I_W} \\ e_{S_W} \\ e_{S_W^2} \\ \vdots \\ e_{S_W^H} \end{bmatrix} \tag{2.50}
$$

Here, the $\boldsymbol{\beta}_{BY}$ and $\boldsymbol{\beta}_{WX}$ are the design matrices in respective levels with the

elements of intercept $\beta_{0.}$ and the regression weight $\beta_{1.}$, and the $\mathbf{e}_B \sim MVN\left(\mathbf{0}, \Theta_{\mathbf{e}_B}\right)$ and

$\mathbf{e}_W \sim MVN\left(\mathbf{0}, \Theta_{\mathbf{e}_W}\right)$ are the random vector of the residual terms of growth factors in two

levels. The cross-correlation of these two level residual terms is assumed to be

uncorrelated.

$$
\begin{aligned}
\mathbf{y}_{ig} &= \boldsymbol{\mu}_B + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B..g} + \boldsymbol{\varepsilon}_{B..g} + \boldsymbol{\mu}_W + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{W.ig} + \boldsymbol{\varepsilon}_{W.ig} \\
&= \boldsymbol{\mu}_B + \boldsymbol{\Lambda}_B \left(\boldsymbol{\beta}_{BY} Y + \mathbf{e}_B \right) + \boldsymbol{\varepsilon}_{B..g} + \boldsymbol{\mu}_W + \boldsymbol{\Lambda}_W \left(\boldsymbol{\beta}_{WX} X + \mathbf{e}_W \right) + \boldsymbol{\varepsilon}_{W.ig}
\end{aligned} \tag{2.51}
$$

And the variance-covariance matrix of $\mathbf{y}_{ig}$ of Equation (2.51) can be also

rewritten as

$$
\begin{aligned}
Cov\left(\mathbf{y}_{ig}\right) &= \mathbf{\Sigma}_B + \mathbf{\Sigma}_W \\
&= \mathbf{\Lambda}_B \mathbf{\Psi}_B \mathbf{\Lambda}_B^{'} + \mathbf{\Theta}_B + \mathbf{\Lambda}_W \mathbf{\Psi}_W \mathbf{\Lambda}_W^{'} + \mathbf{\Theta}_W \\
&= \mathbf{\Lambda}_B Cov\left(\mathbf{\beta}_{BY}\mathbf{Y} + \mathbf{e}_B\right)\mathbf{\Lambda}_B^{'} + \mathbf{\Theta}_B + \mathbf{\Lambda}_W Cov\left(\mathbf{\beta}_{WX}\mathbf{X} + \mathbf{e}_W\right)\mathbf{\Lambda}_W^{'} + \mathbf{\Theta}_W \\
&= \mathbf{\Lambda}_B\mathbf{\beta}_{BY}Cov\left(\mathbf{Y}\right)\mathbf{\beta}_{BY}^{'}\mathbf{\Lambda}_B^{'} + \mathbf{\Lambda}_B\mathbf{\Theta}_{\mathbf{e}_B}\mathbf{\Lambda}_B^{'} + \mathbf{\Theta}_B \\
&\quad + \mathbf{\Lambda}_W\mathbf{\beta}_{WX}Cov\left(\mathbf{X}\right)\mathbf{\beta}_{WX}^{'}\mathbf{\Lambda}_W^{'} + \mathbf{\Lambda}_W\mathbf{\Theta}_{\mathbf{e}_W}\mathbf{\Lambda}_W^{'} + \mathbf{\Theta}_W
\end{aligned}
\tag{2.52}
$$

The between- and within- level variances are assumed to be orthogonal with covariates from respective levels included in the model. The current study aimed to investigate the consequence of failing to model a higher-level structure and the impact of introducing higher-level covariates to the regular and model-based LGCMs with different estimation methods. A Monte Carlo simulation procedure was depicted in the following section.

4.3 Method

A Monte Carlo simulation procedure was conducted to investigate the effect of ignoring the dependent nature of the complex survey data. The simulation considered the following design factors: 3 Model specifications (MLGCM with low-/high-level covariates, LGCM with low-level covariate, and LGCM with low-/high-level covariates), 2 Estimation methods (maximum likelihood estimation (ML), and maximum likelihood estimation with robust standard estimator (MLR)), 3 Cluster sizes (CS=5, 30, 50), and 4 cluster numbers (CN=10, 30, 50, 100) to generate the dependent data. The number of

cluster size and cluster number was chosen based on Maas and Hox (2005) and Kreft

(2006). MLGCM using ML or MLR estimation with the multivariate normally distributed

indicators yielded identical results; thus, only result from MLGCM with MLR estimation

or model-based MLGCM was reported. The combination of model specifications and

estimation methods resulted in 5 different models. In addition to model-based MLGCM

with low- and high-level covariates, the other four models were 1) conditioned regular

LGCM with low-level covariate, 2) design-based MLGCM with low-level covariate, 3)

conditioned regular LGCM with low- and high-level covariates, and 4) design-based

MLGCM with low- and high-level covariates. More information about the five models

was provided in the model specification section. We adopted the Monte Carlo procedure

in Mplus V5.21 (Muthén & Muthén, 2009) for data generation. A total number of 5

(model combinations)$\times 3$ (CS)$\times 4$ (CN) = 60 conditions was examined with 1000

replications for each condition.

4.3.1 Data generation

In this study, we generated a set of complex survey data that was dependent in

nature with a continuous outcome variable for a two-level four-wave growth curve model

(i.e. a three-level model in the multilevel data analysis) as shown in Figure 5, where the

between-level structure had a cluster-level covariate W and the within-level structure had

an individual-level covariate X respectively. We chose 4 waves as the number of repeated

measures based on the review of the multiwave longitudinal studies published in

Developmental Psychology in 2002 by Khoo and colleagues (Khoo, West, Wu, & Kwok,

2006). The four repeated measures were denoted as Y1-Y4. The intercepts of the outcome

variable at the four time points were fixed at zero. The ICC is defined as the ratio between

cluster-level variance and the total variance of a variable (Cohen, Cohen, West, & Aiken,

2003; Muthén & Satorra, 1995; Shrout & Fleiss, 1979). The ICC of the four-wave

measures (e.g. Y1, Y2, Y3, Y4) in the Two-level model were 0.25, 0.27, 0.30, and 0.32,

which were within the range of the commonly found ICCs in educational research (Hox,

2002). The linear growth patterns were modeled in both within- and between-level, i.e. the

factor loadings of the intercept factor were fixed at one and those of the slope factor were

set as 0, 1, 2, and 3 as part of the growth model parameterization. We followed the

notations for parameters used in Duncan, Duncan, and Strycker (2006). The intercepts

were set as 1 for between- and within-level intercept factors (MI_B and MI_W), while the

intercepts of both between- and within-level slope factors (MS_B and MS_W) were set as 0.5.

At the within level, the residual variances of the outcome variables were constrained to be

equal over time and fixed at 0.5. The regression coefficients of intercept (FI_W) and slope

factors (FS_W) on the within-level covariate X were set as 1 and 0.2. The residual variances

of the within intercept factor (DI_W) and of slope factor (DS_W) and covariance (RIS_W) of the

growth factors were set as 1, 0.5, and 0 respectively. At the between level, the residual

variances of the outcome variables were set as zero. The regression coefficients of the

growth factors on the between-level covariate "W" were set to be 0.5 for FI_B on W and 0.2

for FS_B on W. Residual variance of the intercept factor (DI_B) was set as 0.5 and that of the

slope factor (DS_B) was set as 0.2. The covariance between the growth factors (RIS_B) was

set as 0. A similar conditioned MLGCM setting can be referred to in the Mplus manual

(Muthén & Muthén, 1998-2007) .
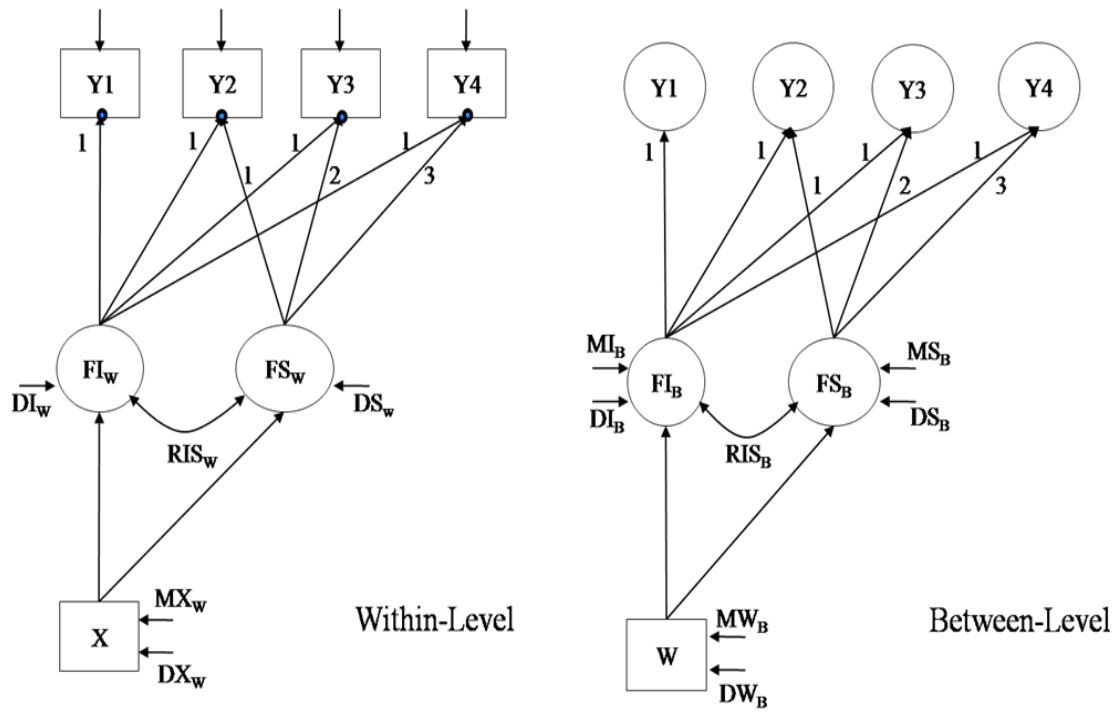
Figure 5 A two-level latent growth curve model with continuous global covariates.

4.3.2 Model specification

In this study, we calculated results for three model specifications using the same

clustered data and compared their results with different estimation methods (i.e. ML or

MLR). Three model specifications were 1) a true conditioned MLGCM with both the

within-level covariate X and between-level covariate W, 2) a one-level LGCM with only

within-level covariate X, and 3) a one-level LGCM with both the within-level covariate X

and between-level covariate W. The three model specifications were illustrated in Figure 5

to Figure 7. The first model specification, the conditioned two-level model, was a

MLGCM using the model-based analytic approach. The default estimator for this

model-based MLGCM analysis is maximum likelihood estimator with robust standard

errors denoted as a MLR estimator in the routine of TYPE=TWOLEVEL in Mplus V5.21

(Muthén & Muthén, 2009). The other two model specifications were run using LGCM with

two distinct types of estimation methods, including 1) maximum likelihood estimator with

robust standard error correction denoted as MLR estimator in the routine of

TYPE=COMPLEX in Mplus in conjunction with the CLUSTER option (i.e. the

design-based analytic approach) and 2) maximum likelihood estimator without robust

standard error correction denoted as ML in the Mplus routine of TYPE=BASIC which is

the regular LGCM analytic approach. For ease of differentiation, we used the following

naming scheme for the five combinations of model specification and model estimation:

1. True Model: The first MLGCM specification with MLR estimator (i.e. the

    model-based MLGCM approach).

2. 1-MLR-X: The LGCM with only within-level covariate X using MLR

    estimator (i.e. a design-based MLGCM approach with only low-level covariate

    X)

3. 1-ML-X: The LGCM with only within-level covariate X using ML estimator

    (i.e., a regular LGCM approach with only low-level covariate X but without

    taking the dependent data information into account)

4. 1-MLR-XW: The LGCM with both the within-level covariate X and

    between-level covariate W using MLR estimator (i.e. a design-based MLGCM

    approach with both high-/low-level covariate W and X)

5. 1-ML-XW: The LGCM with both the within-level covariate X and

    between-level covariate W using ML estimator (i.e., a regular LGCM approach

    with both high-/low-level covariate W and X but without taking dependency

    into account).

Therefore, we had one model-based MLGCM (i.e. the True model, MLGCM), two design-based MLGCM and two regular LGCMs. The 1-ML-X and 1-MLR-X models only had model specification on the within level of the True model, i.e. the between level covariate was neglected. The 1-ML-XW and 1-MLR-XW models had the same specification as the True model in the within level and incorporated the between-level covariate, W. Except for the true model or model-based MLGCM, the other 4 models belong to macro-micro multilevel situations where the dependent variable at the lower level (micro-level) are predicted by variables measured at that lower level or the higher level (macro-level) (Croon & van Veldhoven, 2007) or global or integral variables (Blakely & Woodward, 2000).
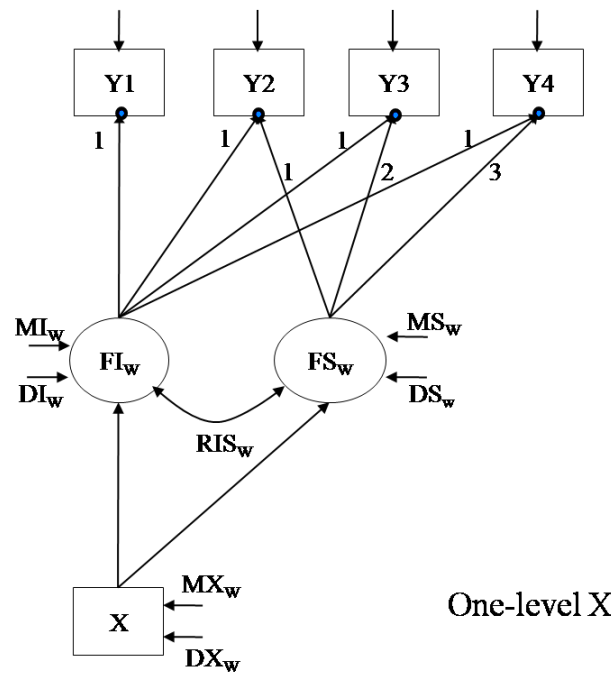
Figure 6. A single-level growth curve model with the individual-level covariate only
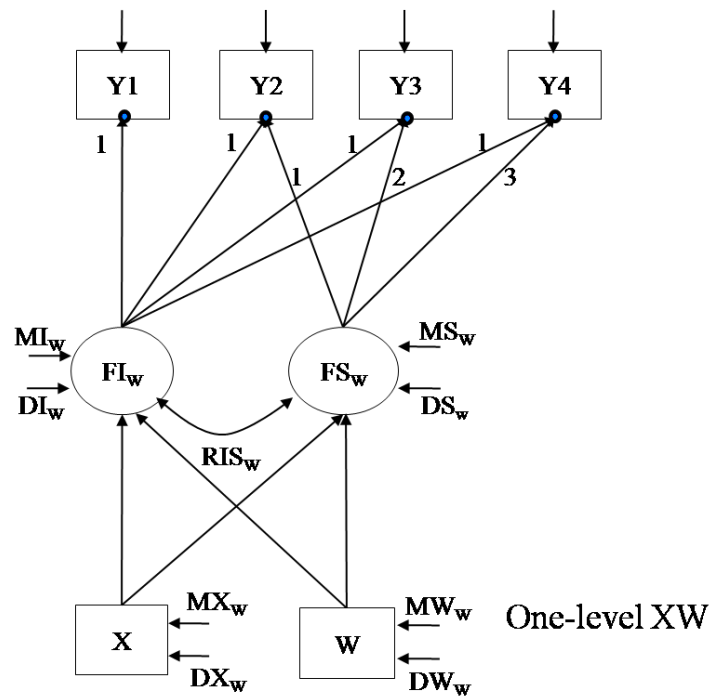
Figure 7. A single-level growth curve model with both the individual-level covariate and cluster-level covariate, (One-level XW model)

4.4 Result

The results are organized so that comparison of the five models with three different

model specifications and two different estimation methods are made with regard to

differences in convergence rate, likelihood ratio test statistic values, model fit indices,

between/within-level regression weights, and between/within-level factor covariance,

residual and mean structure estimates. We also discuss the 95% coverage rate and

empirical power or Type I error rate associated with the estimates of the regression

coefficients, random effects, and the mean structures. The column labeled "sig" is the

empirical power or Type I error rate according to the population value set for the parameter.

If the population value is set as zero, "sig" represents Type I error rate; otherwise, "sig" is

the empirical power to detect the population value.

4.4.1 Convergence rate

The convergence rate equaled 1.0 in all sample size settings of the four 1-level

models and in most of the settings of the True model. In particular, the convergent rate

decreased in the True model when cluster number decreased to 10 (e.g. for cluster

number=10 and cluster size =5, the convergence rate is 0.930).

4.4.2 Likelihood ratio model fit test statistic and model fit indices

The Likelihood ratio test statistic (i.e. χ^2 model fit test statistic) values and model

fit indices showed an overall good fit of the model to the data, shown in Table 10.

Compared to the True model, the χ^2 test statistic values of 1-MLR-X, 1-ML-X,

1-MLR-XW, and 1-ML-XW became smaller due to the decrease of degree of freedom.

Almost all settings across the five models had CFI and TLI equal to 1 except for the

setting at CN=10 and CS=5. All the CFI and TLI values were greater than the convention

cutoff scores (CFI >.95, Hu & Bentler, 1999). RMSEA also revealed a good fit of the

model to the data (RMSEA<.05, Hu & Bentler, 1999) except the setting at CN=10 and

CS=5 for the True model (RMSEA=0.1) due to small cluster number at the between level

and small sample size (n=50). SRMRW and SRMRB were reported for the True model.

All SRMRW and SRMRB values were lower than the traditional cutoff score

(SRMR<.05, Hu and Bentler, 1999). The two design-based MLGCM and two regular

LGCMs reported a single SRMR value. All values also met the conventional rule of

thumb.

Table 10. Model Test Statistic and Fit Indices for Two-level, and One-level MLR Models

		True model						1-MLR-X					1-MLR-XW				
CN	CS	Chi	CFI	TLI	RMSEA	SRMRW	SRMRB	Chi	CFI	TLI	RMSEA	SRMR	Chi	CFI	TLI	RMSEA	SRMR
100	50	17.10	1.00	1.00	0.00	0.00	0.00	10.04	1.00	1.00	0.00	0.00	12.12	1.00	1.00	0.00	0.00
	30	19.90	1.00	1.00	0.01	0.00	0.00	10.01	1.00	1.00	0.01	0.00	12.14	1.00	1.00	0.01	0.00
	5	20.43	1.00	1.00	0.01	0.01	0.01	10.34	1.00	1.00	0.01	0.01	12.50	1.00	1.00	0.01	0.01
50	50	17.30	1.00	1.00	0.00	0.00	0.00	10.18	1.00	1.00	0.01	0.01	12.34	1.00	1.00	0.01	0.00
	30	20.80	1.00	1.00	0.01	0.00	0.01	10.09	1.00	1.00	0.01	0.01	12.33	1.00	1.00	0.01	0.01
	5	21.39	1.00	1.00	0.02	0.01	0.02	10.51	1.00	1.00	0.02	0.02	12.82	1.00	1.00	0.02	0.01
30	50	19.15	1.00	1.00	0.01	0.00	0.01	10.20	1.00	1.00	0.01	0.01	12.51	1.00	1.00	0.01	0.01
	30	21.87	1.00	1.00	0.01	0.00	0.01	10.39	1.00	1.00	0.01	0.01	12.66	1.00	1.00	0.01	0.01
	5	22.78	1.00	1.00	0.03	0.01	0.02	10.64	1.00	1.00	0.02	0.02	13.05	1.00	1.00	0.02	0.02
10	50	19.48	1.00	1.00	0.01	0.01	0.01	10.80	1.00	1.00	0.01	0.01	13.94	1.00	1.00	0.02	0.01
	30	28.13	1.00	1.00	0.03	0.01	0.01	10.87	1.00	1.00	0.02	0.01	13.77	1.00	1.00	0.02	0.01
	5	30.94	0.97	0.97	0.10	0.02	0.04	11.66	0.99	1.00	0.05	0.04	14.91	0.99	0.99	0.06	0.03

Note. CN=Cluster Number; CS=Cluster Size; Chi: Yuan-Bentler T2* test statistic (Muthén & Muthén, 2009); CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Squared error of approximation; SRMRB= between-level standardized root mean squared residual; SRMRW= within-level standardized root mean squared residual; SRMR= standardized root mean squared residual.

Table 10 continued.

CN	CS	1-ML-X					1-ML-XW				
		Chi	CFI	TLI	RMSEA	SRMR	Chi	CFI	TLI	RMSEA	SRMR
	50	10.02	1.00	1.00	0.00	0.00	12.00	1.00	1.00	0.00	0.00
100	30	9.95	1.00	1.00	0.00	0.00	11.98	1.00	1.00	0.00	0.00
	5	10.27	1.00	1.00	0.01	0.01	12.35	1.00	1.00	0.01	0.01
	50	10.10	1.00	1.00	0.01	0.01	12.06	1.00	1.00	0.01	0.00
50	30	10.01	1.00	1.00	0.01	0.01	12.04	1.00	1.00	0.01	0.01
	5	10.38	1.00	1.00	0.02	0.02	12.49	1.00	1.00	0.02	0.01
	50	10.06	1.00	1.00	0.01	0.01	12.07	1.00	1.00	0.01	0.01
30	30	10.21	1.00	1.00	0.01	0.01	12.16	1.00	1.00	0.01	0.01
	5	10.41	1.00	1.00	0.02	0.02	12.49	1.00	1.00	0.02	0.02
	50	10.19	1.00	1.00	0.01	0.01	12.24	1.00	1.00	0.01	0.01
10	30	10.23	1.00	1.00	0.02	0.01	12.17	1.00	1.00	0.02	0.01
	5	10.86	0.99	1.00	0.04	0.04	13.14	0.99	1.00	0.04	0.03

Note. CN=Cluster Number; CS=Cluster Size; Chi: Yuan-Bentler T2* test statistic (Muthén & Muthén, 2009); CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Squared error of approximation; SRMRB= between-level standardized root mean squared residual; SRMRW= within-level standardized root mean squared residual; SRMR= standardized root mean squared residual;

4.4.3 Regression weight estimates

Regression coefficients for FI_W and FS_W on lower-level covariate X were reported

for all the models, while regression coefficients for FI_B and FS_B on W or FI_W and FS_W on

higher-level covariate W were reported for the True model, 1-ML-XW, and

1-MLR-XWmodels only, because covariate W was not added in these two models. The

estimates of the regression weights, their corresponding standard errors, 95% coverage

rate, and empirical power are presented in Table 11. Generally speaking, the relative bias

of the fixed effect estimates remained unbiased across the five models. The standard error,

95% coverage rate, and empirical power were similar across the True model, 1-MLR-X,

1-MLR-XW, and 1-ML-X models but the 1-ML-XW model showed a different pattern for

FI_W on W and FS_W on W. After the cluster-level predictor was added in the 1-ML-XW

model the fixed effect estimates on W remained unbiased but the corresponding standard errors were underestimated, the 95% coverage rate shrunken, and the empirical power inflated at smaller cluster number settings.

What is also worth our attention was the comparison of regression coefficients between FI_B on higher-level covariate W and FS_B on higher-level covariate W in the True model and those for FI_W on higher-level covariate W and FS_W on W in the 1-MLR XW model. Contrary to the 1-ML-XW model, when higher-level covariate W was incorporated in the within level to predict growth factors FI_W and FS_W in the 1-MLR-XW model, it produced consistent but less efficient regression coefficient estimates with congruent 95% coverage rate and empirical power as those in the True model. However, in the small cluster number setting (e.g. CN=10), the 95% CI coverage rate and empirical power diminished around 10% to 20% as a result of power issue.

Table 11.   Regression Coefficient Estimates between Covariates and Growth Factors.

	CN	CS	FI_W on X				FS_W on X				FI_B on W / FI_W on W				FS_B on W / FS_W on W			
			Est.	SE	95%	Sig.	Est.	SE	95%	Sig.	Est.	SE	95%	Sig.	Est.	SE	95%	Sig.
True model	100	50	1.00	0.02	0.94	1.00	0.20	0.01	0.95	1.00	0.50	0.07	0.93	1.00	0.20	0.05	0.92	0.98
		30	1.00	0.02	0.96	1.00	0.20	0.01	0.95	1.00	0.50	0.07	0.94	1.00	0.20	0.05	0.94	0.98
		5	1.00	0.06	0.94	1.00	0.20	0.04	0.94	1.00	0.50	0.09	0.94	1.00	0.20	0.06	0.92	0.93
	50	50	1.00	0.02	0.94	1.00	0.20	0.02	0.95	1.00	0.50	0.10	0.93	1.00	0.20	0.06	0.92	0.85
		30	1.00	0.03	0.95	1.00	0.20	0.02	0.94	1.00	0.51	0.10	0.93	1.00	0.20	0.06	0.93	0.86
		5	1.00	0.08	0.94	1.00	0.20	0.05	0.94	0.97	0.50	0.12	0.91	0.97	0.20	0.08	0.90	0.71
	30	50	1.00	0.03	0.94	1.00	0.20	0.02	0.93	1.00	0.49	0.13	0.90	0.95	0.20	0.08	0.90	0.69
		30	1.00	0.04	0.95	1.00	0.20	0.03	0.93	1.00	0.51	0.13	0.92	0.94	0.20	0.08	0.89	0.67
		5	1.00	0.10	0.93	1.00	0.20	0.07	0.94	0.85	0.50	0.15	0.91	0.85	0.19	0.10	0.90	0.52
	10	50	1.00	0.05	0.91	1.00	0.20	0.03	0.89	1.00	0.49	0.20	0.82	0.67	0.19	0.12	0.80	0.43
		30	1.00	0.06	0.90	1.00	0.20	0.04	0.91	0.99	0.51	0.21	0.80	0.67	0.20	0.13	0.82	0.44
		5	1.01	0.16	0.89	1.00	0.19	0.11	0.89	0.46	0.50	0.25	0.82	0.56	0.19	0.16	0.83	0.33
1-MLR-X	100	50	1.00	0.02	0.95	1.00	0.20	0.01	0.96	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.03	0.95	1.00	0.20	0.02	0.93	1.00	.	.	.	.	.	.	.	.
		5	1.00	0.06	0.95	1.00	0.20	0.04	0.95	1.00	.	.	.	.	.	.	.	.
	50	50	1.00	0.03	0.94	1.00	0.20	0.02	0.94	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.04	0.94	1.00	0.20	0.02	0.93	1.00	.	.	.	.	.	.	.	.
		5	1.00	0.09	0.95	1.00	0.20	0.06	0.95	0.94	.	.	.	.	.	.	.	.
	30	50	1.00	0.04	0.94	1.00	0.20	0.02	0.94	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.05	0.94	1.00	0.20	0.03	0.94	1.00	.	.	.	.	.	.	.	.
		5	1.00	0.12	0.94	1.00	0.20	0.07	0.95	0.79	.	.	.	.	.	.	.	.
	10	50	1.00	0.06	0.92	1.00	0.20	0.04	0.91	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.08	0.91	1.00	0.20	0.05	0.92	0.97	.	.	.	.	.	.	.	.
		5	1.01	0.19	0.91	1.00	0.19	0.12	0.91	0.39	.	.	.	.	.	.	.	.
1-MLR-XW	100	50	1.00	0.02	0.95	1.00	0.20	0.01	0.96	1.00	0.50	0.07	0.93	1.00	0.20	0.05	0.92	0.98
		30	1.00	0.02	0.95	1.00	0.20	0.02	0.93	1.00	0.50	0.07	0.94	1.00	0.20	0.05	0.94	0.98
		5	1.00	0.06	0.96	1.00	0.20	0.04	0.95	1.00	0.50	0.09	0.94	1.00	0.20	0.06	0.92	0.93
	50	50	1.00	0.03	0.93	1.00	0.20	0.02	0.94	1.00	0.50	0.10	0.93	1.00	0.20	0.06	0.92	0.85
		30	1.00	0.03	0.94	1.00	0.20	0.02	0.94	1.00	0.51	0.10	0.93	1.00	0.20	0.06	0.93	0.86
		5	1.00	0.08	0.94	1.00	0.20	0.06	0.94	0.94	0.50	0.12	0.92	0.97	0.20	0.08	0.91	0.71
	30	50	1.00	0.03	0.95	1.00	0.20	0.02	0.94	1.00	0.49	0.13	0.91	0.94	0.20	0.08	0.91	0.68
		30	1.00	0.04	0.95	1.00	0.20	0.03	0.94	1.00	0.51	0.13	0.93	0.94	0.20	0.08	0.90	0.66
		5	1.00	0.11	0.93	1.00	0.20	0.07	0.94	0.80	0.50	0.16	0.91	0.84	0.19	0.10	0.90	0.52
	10	50	1.00	0.06	0.92	1.00	0.20	0.04	0.90	1.00	0.49	0.21	0.84	0.65	0.19	0.13	0.83	0.40
		30	1.00	0.07	0.92	1.00	0.20	0.05	0.92	0.97	0.51	0.22	0.82	0.63	0.20	0.13	0.84	0.41
		5	1.01	0.17	0.89	1.00	0.19	0.12	0.90	0.41	0.51	0.26	0.83	0.56	0.20	0.17	0.84	0.32

Note: Population value was set as follows: FI_W on X =1, FS_W on X =.20, FI_B on W =0.5 FS_B on W =0.2, FI_W on W =0.5, and FS_W on W =0.2. Est =Parameter estimate; SE= standard error of parameter estimate; 95% = the 95% confidence interval coverage rate; Sig = proportion of replication of parameter estimates which is statistically significantly different from zero at the .05 level.

Table 11 continued.

	CN	CS	FI$_W$ on X				FS$_W$ on X				FI$_B$ on W / FI$_W$ on W				FS$_B$ on W / FS$_W$ on W			
			Est.	SE	95%	Sig.	Est.	SE	95%	Sig.	Est.	SE	95%	Sig.	Est.	SE	95%	Sig.
1-ML-X	100	50	1.00	0.02	0.95	1.00	0.20	0.01	0.96	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.03	0.95	1.00	0.20	0.02	0.94	1.00	.	.	.	.	.	.	.	.
		5	1.00	0.06	0.96	1.00	0.20	0.04	0.95	1.00	.	.	.	.	.	.	.	.
	50	50	1.00	0.03	0.94	1.00	0.20	0.02	0.94	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.04	0.95	1.00	0.20	0.02	0.94	1.00	.	.	.	.	.	.	.	.
		5	1.00	0.09	0.96	1.00	0.20	0.06	0.95	0.94	.	.	.	.	.	.	.	.
	30	50	1.00	0.04	0.96	1.00	0.20	0.02	0.95	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.05	0.95	1.00	0.20	0.03	0.95	1.00	.	.	.	.	.	.	.	.
		5	1.00	0.12	0.96	1.00	0.20	0.07	0.96	0.78	.	.	.	.	.	.	.	.
	10	50	1.00	0.06	0.95	1.00	0.20	0.04	0.94	1.00	.	.	.	.	.	.	.	.
		30	1.00	0.08	0.95	1.00	0.20	0.05	0.95	0.96	.	.	.	.	.	.	.	.
		5	1.01	0.20	0.94	1.00	0.19	0.13	0.94	0.34	.	.	.	.	.	.	.	.
1-ML-XW	100	50	1.00	0.02	0.95	1.00	0.20	0.01	0.96	1.00	0.50	0.02	0.40	1.00	0.20	0.01	0.42	1.00
		30	1.00	0.02	0.95	1.00	0.20	0.02	0.94	1.00	0.50	0.03	0.49	1.00	0.20	0.02	0.50	1.00
		5	1.00	0.06	0.96	1.00	0.20	0.04	0.95	1.00	0.50	0.06	0.82	1.00	0.20	0.04	0.82	0.98
	50	50	1.00	0.03	0.94	1.00	0.20	0.02	0.95	1.00	0.50	0.03	0.42	1.00	0.20	0.02	0.42	0.99
		30	1.00	0.03	0.96	1.00	0.20	0.02	0.94	1.00	0.51	0.04	0.48	1.00	0.20	0.02	0.53	0.99
		5	1.00	0.09	0.95	1.00	0.20	0.06	0.95	0.94	0.50	0.09	0.80	0.99	0.20	0.06	0.82	0.84
	30	50	1.00	0.03	0.96	1.00	0.20	0.02	0.95	1.00	0.49	0.04	0.39	1.00	0.20	0.02	0.41	0.96
		30	1.00	0.04	0.96	1.00	0.20	0.03	0.95	1.00	0.51	0.05	0.46	1.00	0.20	0.03	0.50	0.94
		5	1.00	0.11	0.94	1.00	0.20	0.07	0.95	0.79	0.50	0.11	0.79	0.94	0.19	0.08	0.81	0.67
	10	50	1.00	0.06	0.95	1.00	0.20	0.04	0.94	1.00	0.49	0.07	0.39	0.92	0.19	0.04	0.40	0.81
		30	1.00	0.08	0.95	1.00	0.20	0.05	0.95	0.97	0.51	0.09	0.48	0.91	0.20	0.06	0.51	0.76
		5	1.01	0.19	0.93	1.00	0.19	0.12	0.94	0.37	0.51	0.21	0.79	0.66	0.20	0.14	0.81	0.40

Note: Population value was set as follows: FI$_W$ on X =1, FS$_W$ on X =.20, FI$_B$ on W =0.5 FS$_B$ on W =0.2, FI$_W$ on W =0.5, and FS$_W$ on W =0.2. Est =Parameter estimate; SE= standard error of parameter estimate; 95% = the 95% confidence interval coverage rate; Sig = proportion of replication of parameter estimates which is statistically significantly different from zero at the .05 level.

4.4.4 Random effect estimates

The random effect estimates of growth factors are shown in Table 12. For the

covariance between growth factors, the covariance estimates of the True model (mean

RIS$_w$ =0.0017), the 1-MLR-XW model (mean RIS$_w$ =0.0008), and 1-ML-XW model (mean

$RIS_w = 0.0008$) were much closer to the population parameter, which was set as zero, than

those from the 1-MLR-X model (mean $RIS_w = 0.0975$) and 1-ML-X (mean $RIS_w = 0.0975$).

The True model and the 1-MLR-XW model had more consistent 95% confidence

interval (95% CI) coverage rate than that of the 1-MLR-X model. Saying that the 95% CI

was consistent implied that the empirical 95% CI coverage rate was close to the theoretical

value, 0.95. The True model and the 1-MLR-XW Model also had more accurate

percentage of significant coefficients, which is the empirical rate of parameter estimates

being statistically significant or Type I error rate, the probability of rejecting the null

hypothesis when the population value is set as zero (i.e. when null hypothesis is true.)

(Muthén & Muthén, 1998-2007). The population value of covariance between FI_W and

FS_W was set as zero. A more accurate percentage of significant coefficients under this

setting implied a small empirical Type I error rate to detect a significant parameter estimate

that was different from zero. The True model (mean Type I error rate =.0675) and the

1-MLR-XW model (mean Type I error rate =.0683) had relatively small empirical Type I

error rate in detecting a significant factor covariance estimate compared to the 1-MLR-X

model (mean Type I error rate =.2608). Compared with the MLR models, the ML models

had underestimated SEs for the covariance estimates, inconsistent (lowered) 95% coverage

rate, and inflated Type I error rate (mean Type I error rate =.5292 for the 1-ML-X model

and .2117 for the 1-ML-XW model).

The inconsistent and inefficient covariance estimates combined with the

inconsistent 95% CI and inflated type I error rate or inflated empirical power in the

1-MLR-X, 1-ML-X, and 1-ML-XW models provided erroneous statistical inference,

especially in large sample size settings(e.g. for CN(CS)=100(50) in the 1-ML-X model, 95%

CI coverage rate was 0.08, and rate of significant estimate was 0.92). After the cluster-level

predictor was introduced in the 1-MLR-XW model, the covariance estimates were

consistent with those produced by the True model. As a result, the 1-MLR-XW model

provided more reliable statistical inferences by giving a more conservative percentage of

significant coefficients when the dependent nature of the data was neglected, especially as

sample size became large (e.g. for CN(CS)=100(50) in the 1-MLR-XW model, 95% CI

coverage rate was 0.95, and rate of significant estimate was 0.05). Though the cluster-level

predictor W was also included in the 1-ML-XW model, the ML estimator without robust

standard error correction underestimated SEs for the covariance estimates, produced

inconsistent (lowered) 95% coverage rate, and resulted in inflated Type I error rate (e.g. for

CN(CS)=100(50) in the 1-ML-XW model, 95% CI coverage rate was 0.68, and rate of

significant estimate was 0.32).

For the estimates of growth factor residual variances, the estimated residual

variance in the 1-MLR-X and 1-ML-X models were larger than those in the 1-MLR-XW

and 1-ML-XW models, where the relationship between W and growth factors were

investigated (e.g. for 100(50), DI_W and DS_W was 1.74 and 0.74 in the 1-MLR-X and

1-ML-X models while DI_W and DS_W was 1.49 and 0.70 in the 1-MLR-XW and 1-ML-XW

models). The estimates of residual variance in the 1-MLR-XW and 1-ML-XW models can

be simply viewed as the summation of the residual variance component estimates from the

within- and between-levels in True model (e.g. for 100(50) in the True model, DI_W and

DS_W were 1.00 and 0.50 in and DI_B and DS_B were 0.49 and 0.20 in between-level).

Compared with the 1-MLR models, the 1-ML models underestimated SEs for the factor

residual variances; however, due to the redistribution of variance components both the

1-MLR and the 1-ML models had inconsistent 95% coverage rate and inflated empirical

power.

Table 12. Covariance and Residual Variance estimates of Growth Factors.

	CN	CS	RIS_W Est	SE	95%	Sig	DI_W Est	SE	95%	Sig	DS_W Est	SE	95%	Sig	DI_B Est	SE	95%	Sig	DS_B Est	SE	95%	Sig
True model	100	50	0.00	0.01	0.96	0.04	1.00	0.03	0.94	1.00	0.50	0.01	0.95	1.00	0.49	0.07	0.93	1.00	0.20	0.03	0.92	1.00
		30	0.00	0.02	0.94	0.06	1.00	0.04	0.94	1.00	0.50	0.02	0.94	1.00	0.49	0.07	0.93	1.00	0.20	0.03	0.91	1.00
		5	0.00	0.05	0.94	0.06	1.00	0.10	0.94	1.00	0.50	0.04	0.95	1.00	0.49	0.11	0.92	1.00	0.19	0.04	0.90	1.00
	50	50	0.00	0.02	0.95	0.05	1.00	0.04	0.94	1.00	0.50	0.02	0.95	1.00	0.48	0.10	0.89	1.00	0.19	0.04	0.89	1.00
		30	0.00	0.02	0.94	0.06	1.00	0.05	0.93	1.00	0.50	0.02	0.94	1.00	0.47	0.10	0.89	1.00	0.19	0.04	0.87	1.00
		5	0.00	0.06	0.92	0.08	1.00	0.13	0.95	1.00	0.50	0.06	0.93	1.00	0.47	0.14	0.88	0.98	0.19	0.06	0.87	0.94
	30	50	0.00	0.02	0.93	0.07	1.00	0.05	0.92	1.00	0.50	0.02	0.94	1.00	0.47	0.12	0.83	1.00	0.19	0.05	0.85	1.00
		30	0.00	0.03	0.94	0.06	1.00	0.06	0.94	1.00	0.50	0.03	0.93	1.00	0.46	0.12	0.85	1.00	0.18	0.05	0.83	1.00
		5	0.01	0.08	0.93	0.07	0.98	0.17	0.91	1.00	0.50	0.07	0.92	1.00	0.46	0.18	0.85	0.80	0.18	0.07	0.84	0.71
	10	50	-0.00	0.04	0.91	0.09	1.00	0.08	0.88	1.00	0.50	0.04	0.89	1.00	0.39	0.15	0.65	0.84	0.16	0.06	0.65	0.83
		30	0.00	0.05	0.92	0.08	1.00	0.10	0.89	1.00	0.50	0.05	0.91	1.00	0.40	0.16	0.69	0.78	0.15	0.06	0.64	0.76
		5	0.01	0.13	0.91	0.09	0.93	0.26	0.84	0.98	0.47	0.12	0.86	1.00	0.38	0.27	0.75	0.22	0.15	0.12	0.79	0.21
1-MLR-X	100	50	0.10	0.05	0.43	0.58	1.74	0.11	0.00	1.00	0.74	0.04	0.00	1.00								
		30	0.10	0.05	0.44	0.56	1.74	0.11	0.00	1.00	0.74	0.04	0.00	1.00								
		5	0.10	0.07	0.72	0.28	1.73	0.16	0.00	1.00	0.73	0.06	0.01	1.00								
	50	50	0.10	0.06	0.69	0.31	1.74	0.15	0.00	1.00	0.74	0.05	0.00	1.00								
		30	0.10	0.07	0.68	0.32	1.74	0.16	0.00	1.00	0.73	0.05	0.00	1.00								
		5	0.10	0.10	0.82	0.18	1.73	0.22	0.05	1.00	0.73	0.08	0.18	1.00								
	30	50	0.10	0.08	0.78	0.22	1.72	0.19	0.00	1.00	0.73	0.06	0.00	1.00								
		30	0.10	0.09	0.81	0.19	1.73	0.20	0.01	1.00	0.73	0.07	0.02	1.00								
		5	0.10	0.12	0.86	0.14	1.71	0.28	0.26	1.00	0.72	0.11	0.44	1.00								
	10	50	0.09	0.12	0.89	0.11	1.65	0.28	0.29	1.00	0.71	0.09	0.40	1.00								
		30	0.09	0.13	0.86	0.14	1.68	0.31	0.41	1.00	0.71	0.10	0.48	1.00								
		5	0.09	0.19	0.90	0.10	1.60	0.43	0.78	0.99	0.69	0.16	0.84	1.00								
1-MLR-XW	100	50	0.00	0.04	0.95	0.05	1.49	0.08	0.00	1.00	0.70	0.03	0.00	1.00								
		30	0.00	0.04	0.96	0.05	1.49	0.08	0.00	1.00	0.70	0.03	0.00	1.00								
		5	-0.00	0.06	0.94	0.07	1.48	0.13	0.02	1.00	0.69	0.06	0.04	1.00								
	50	50	0.00	0.05	0.94	0.06	1.48	0.11	0.00	1.00	0.69	0.04	0.00	1.00								
		30	0.00	0.05	0.95	0.05	1.47	0.11	0.00	1.00	0.69	0.05	0.00	1.00								
		5	0.00	0.08	0.95	0.05	1.47	0.18	0.23	1.00	0.68	0.08	0.31	1.00								
	30	50	0.00	0.06	0.93	0.07	1.46	0.13	0.01	1.00	0.68	0.05	0.01	1.00								
		30	0.00	0.06	0.95	0.05	1.46	0.14	0.02	1.00	0.68	0.06	0.06	1.00								
		5	0.01	0.11	0.93	0.07	1.44	0.23	0.53	1.00	0.67	0.10	0.59	1.00								
	10	50	0.00	0.09	0.89	0.11	1.39	0.19	0.46	1.00	0.65	0.08	0.53	1.00								
		30	0.00	0.10	0.90	0.10	1.39	0.21	0.55	1.00	0.65	0.08	0.59	1.00								
		5	0.00	0.16	0.91	0.09	1.30	0.34	0.88	1.00	0.62	0.15	0.89	1.00								

Note. Population value was set as: Covariance between growth factors (RIS_W) = 0; Residual variance of Intercept factor in within level (DI_W) =1 and between level (DI_B) = 0.5; Residual variance of Slope factor in within level (DS_W) =0.5 and between level (DS_B) = 0.2; Est =Parameter estimate; SE= standard error of parameter estimate; 95% = the 95% confidence interval coverage rate; Sig = proportion of replication of parameter estimates which is statistically significantly different from zero at the .05 level.

Table 12 Continued.

	CN	CS	RIS$_W$				DI$_W$				DS$_W$				DI$_B$				DS$_B$			
			Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig
1-ML-X	100	50	0.10	0.02	0.08	0.92	1.74	0.04	0.00	1.00	0.74	0.02	0.00	1.00								
		30	0.10	0.02	0.13	0.87	1.74	0.05	0.00	1.00	0.74	0.02	0.00	1.00								
		5	0.10	0.06	0.61	0.39	1.73	0.13	0.00	1.00	0.73	0.05	0.01	1.00								
	50	50	0.10	0.03	0.22	0.78	1.74	0.06	0.00	1.00	0.74	0.02	0.00	1.00								
		30	0.10	0.03	0.30	0.70	1.74	0.08	0.00	1.00	0.73	0.03	0.00	1.00								
		5	0.10	0.08	0.74	0.26	1.73	0.19	0.03	1.00	0.73	0.07	0.13	1.00								
	30	50	0.10	0.03	0.35	0.65	1.72	0.08	0.00	1.00	0.73	0.03	0.00	1.00								
		30	0.10	0.04	0.42	0.58	1.73	0.10	0.00	1.00	0.73	0.04	0.01	1.00								
		5	0.10	0.11	0.79	0.21	1.71	0.24	0.15	1.00	0.72	0.10	0.37	1.00								
	10	50	0.09	0.06	0.52	0.48	1.65	0.13	0.06	1.00	0.71	0.05	0.12	1.00								
		30	0.09	0.07	0.61	0.39	1.68	0.17	0.09	1.00	0.71	0.07	0.22	1.00								
		5	0.09	0.18	0.88	0.12	1.60	0.39	0.72	1.00	0.69	0.16	0.85	1.00								
1-ML-XW	100	50	0.00	0.02	0.68	0.32	1.49	0.04	0.00	1.00	0.70	0.02	0.00	1.00								
		30	0.00	0.02	0.76	0.24	1.49	0.05	0.00	1.00	0.70	0.02	0.00	1.00								
		5	-0.00	0.05	0.91	0.09	1.48	0.12	0.01	1.00	0.69	0.05	0.03	1.00								
	50	50	0.00	0.02	0.67	0.33	1.48	0.05	0.00	1.00	0.69	0.02	0.00	1.00								
		30	0.00	0.03	0.77	0.23	1.47	0.07	0.00	1.00	0.69	0.03	0.00	1.00								
		5	0.00	0.08	0.92	0.08	1.47	0.16	0.17	1.00	0.68	0.07	0.26	1.00								
	30	50	0.00	0.03	0.67	0.34	1.46	0.07	0.00	1.00	0.68	0.03	0.00	1.00								
		30	0.00	0.04	0.79	0.21	1.46	0.09	0.00	1.00	0.68	0.04	0.02	1.00								
		5	0.01	0.10	0.91	0.09	1.44	0.21	0.46	1.00	0.67	0.09	0.53	1.00								
	10	50	0.00	0.05	0.69	0.31	1.39	0.11	0.19	1.00	0.65	0.05	0.25	1.00								
		30	0.00	0.07	0.77	0.23	1.39	0.14	0.32	1.00	0.65	0.06	0.39	1.00								
		5	0.00	0.16	0.93	0.07	1.30	0.33	0.89	1.00	0.62	0.14	0.92	1.00								

Note. Population value was set as: Covariance between growth factors (RIS$_W$) = 0; Residual variance of Intercept factor in within level (DI$_W$) =1 and between level (DI$_B$) = 0.5; Residual variance of Slope factor in within level (DS$_W$) =0.5 and between level (DS$_B$) = 0.2; Est =Parameter estimate; SE= standard error of parameter estimate; 95% = the 95% confidence interval coverage rate; Sig = proportion of replication of parameter estimates which is statistically significantly different from zero at the .05 level.

4.4.5 Mean structure estimates

The result of mean structure for the models is shown in Table 13. The True model

had consistent and efficient growth factors mean estimates. However, in the small cluster

number setting (e.g. CN=10), the 95% CI coverage rate and empirical power diminished as

a result of power issue. The same power issue occurred in the 1-MLR-X and 1-MLR-XW

models. Compared with the True model, the 1-MLR-X model had slightly inflated

standard errors on the mean structure estimates; as a result, it had less power than the True

model in detecting the statistical significance of parameter estimates in the small cluster

number settings (e.g. CN=10). By introducing the cluster-level covariate W, the

1-MLR-XW model generated consistent and efficient parameter estimates, 95% CI

coverage rate, and statistical power congruent to the results of the True model across all

sample size settings. Despite the slightly inflated standard errors for the 1-MLR-X model,

the 1-MLR-X and 1-MLR-XW models had negligible differences in the mean structure

results. Compared with the True model and the two 1-MLR models, the 1-ML models

underestimated the SEs for the growth factor mean structures, produced inconsistent 95%

coverage rate, and had inflated empirical power at small cluster number settings.

Table 13. Mean Structure estimates of Growth Factors.

		True model								1-MLR-X								1-MLR-XW							
		MI_B				MS_B				MI_W				MS_W				MI_W				MS_W			
CN	CS	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig
100	50	1.00	0.07	0.95	1.00	0.50	0.05	0.95	1.00	1.00	0.09	0.95	1.00	0.50	0.05	0.95	1.00	1.00	0.07	0.95	1.00	0.50	0.05	0.95	1.00
	30	1.00	0.07	0.95	1.00	0.50	0.05	0.94	1.00	1.00	0.09	0.94	1.00	0.50	0.05	0.95	1.00	1.00	0.07	0.95	1.00	0.50	0.05	0.94	1.00
	5	1.00	0.09	0.95	1.00	0.50	0.06	0.95	1.00	1.00	0.10	0.96	1.00	0.50	0.06	0.96	1.00	1.00	0.09	0.95	1.00	0.50	0.06	0.95	1.00
50	50	1.00	0.10	0.93	1.00	0.50	0.06	0.94	1.00	1.00	0.12	0.94	1.00	0.50	0.07	0.95	1.00	1.00	0.10	0.93	1.00	0.50	0.06	0.94	1.00
	30	1.00	0.10	0.95	1.00	0.50	0.07	0.94	1.00	1.00	0.13	0.95	1.00	0.50	0.07	0.94	1.00	1.00	0.10	0.95	1.00	0.50	0.07	0.94	1.00
	5	1.00	0.12	0.93	1.00	0.50	0.08	0.94	1.00	1.00	0.14	0.94	1.00	0.50	0.08	0.94	1.00	1.00	0.12	0.94	1.00	0.50	0.08	0.95	1.00
30	50	1.00	0.13	0.93	1.00	0.50	0.08	0.93	1.00	1.01	0.16	0.94	1.00	0.50	0.09	0.94	1.00	1.00	0.13	0.93	1.00	0.50	0.08	0.93	1.00
	30	1.00	0.13	0.93	1.00	0.50	0.08	0.93	1.00	1.00	0.16	0.96	1.00	0.50	0.09	0.94	1.00	1.00	0.13	0.94	1.00	0.50	0.08	0.93	1.00
	5	1.00	0.16	0.92	1.00	0.50	0.10	0.93	1.00	1.01	0.18	0.94	1.00	0.50	0.11	0.93	0.99	1.00	0.16	0.93	1.00	0.50	0.10	0.93	1.00
10	50	1.00	0.21	0.86	0.98	0.50	0.13	0.87	0.91	1.01	0.27	0.90	0.94	0.50	0.15	0.91	0.88	1.00	0.22	0.88	0.98	0.50	0.14	0.89	0.89
	30	1.00	0.21	0.88	0.99	0.50	0.13	0.87	0.91	1.01	0.28	0.92	0.94	0.50	0.16	0.92	0.85	1.00	0.23	0.89	0.98	0.50	0.14	0.89	0.90
	5	1.00	0.25	0.88	0.93	0.51	0.17	0.88	0.80	1.01	0.31	0.92	0.87	0.51	0.18	0.91	0.77	1.00	0.26	0.89	0.92	0.51	0.17	0.89	0.79

Note. Intercept of between-level Intercept factor (MI_B) = Intercept of within-level Intercept factor (MI_W) = 1; Intercept of between-level slope factor (MS_B) = Intercept of within-level Slope factor (MS_W)=0.5. Est =Parameter estimate; SE= standard error of parameter estimate; 95% = the 95% confidence interval coverage rate; Sig = proportion of replication of parameter estimates which is statistically significantly different from zero at alpha= .05 level.

Table 13 continued.

CN	CS	1-ML-X								1-ML-XW							
		MI_W				MS_W				MI_W				MS_W			
		Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig	Est	SE	95%	Sig
100	50	1.00	0.02	0.32	1.00	0.50	0.01	0.37	1.00	1.00	0.02	0.37	1.00	0.50	0.01	0.42	1.00
	30	1.00	0.03	0.45	1.00	0.50	0.02	0.46	1.00	1.00	0.02	0.49	1.00	0.50	0.02	0.49	1.00
	5	1.00	0.06	0.80	1.00	0.50	0.04	0.80	1.00	1.00	0.06	0.84	1.00	0.50	0.04	0.83	1.00
50	50	1.00	0.03	0.35	1.00	0.50	0.02	0.37	1.00	1.00	0.03	0.40	1.00	0.50	0.02	0.40	1.00
	30	1.00	0.04	0.42	1.00	0.50	0.02	0.47	1.00	1.00	0.04	0.48	1.00	0.50	0.02	0.50	1.00
	5	1.00	0.09	0.79	1.00	0.50	0.06	0.82	1.00	1.00	0.09	0.81	1.00	0.50	0.06	0.82	1.00
30	50	1.01	0.04	0.36	1.00	0.50	0.02	0.37	1.00	1.00	0.04	0.39	1.00	0.50	0.02	0.40	1.00
	30	1.00	0.05	0.45	1.00	0.50	0.03	0.47	1.00	1.00	0.05	0.48	1.00	0.50	0.03	0.49	1.00
	5	1.01	0.12	0.78	1.00	0.50	0.07	0.81	1.00	1.00	0.11	0.82	1.00	0.50	0.07	0.84	1.00
10	50	1.01	0.06	0.35	1.00	0.50	0.04	0.36	1.00	1.00	0.06	0.39	1.00	0.50	0.04	0.36	1.00
	30	1.01	0.08	0.44	1.00	0.50	0.05	0.47	0.99	1.00	0.08	0.49	1.00	0.50	0.05	0.49	0.99
	5	1.01	0.20	0.77	0.97	0.51	0.13	0.79	0.91	1.00	0.19	0.79	0.97	0.51	0.13	0.80	0.91

Note. Intercept of between-level Intercept factor (MI_B) = Intercept of within-level Intercept factor (MI_W) = 1; Intercept of between-level slope factor (MS_B) = Intercept of within-level Slope factor (MS_W)=0.5. Est =Parameter estimate; SE= standard error of parameter estimate; 95% = the 95% confidence interval coverage rate; Sig = proportion of replication of parameter estimates which is statistically significantly different from zero at alpha= .05 level.

4.5 Power Analysis

Based on the simulation results, the True model and the 1-MLR-XW model

performed equally well on the criterion variables and outperformed the other three models.

In order to examine whether the Type I error rates are also under control for the test of fixed

effects, additional Monte Carlo simulations were conducted to compute the empirical Type

I error rates by setting the fixed effect estimates equal zero across all the sample size

settings for the True model and the 1-MLR-XW model. The empirical Type I error rates

were considered biased if the they fell out of Bradley's (.5α, 1.5 α) liberal definition of

robustness (1978). According to the simulation results in the True model, the mean Type I

error rates at CN=100, 50, 30, and 10 were .041, .058, .057, and .104 for FI_W on

X, .053, .058, .048, and .081 for FS_W on X, .053, .067, .078, .172, for FI_B on W,

and .065, .072, .093, and .173 for FS_B on W. In the 1-MLR-XW model, the mean Type I

error rates at CN=100, 50, 30, and 10 were .045, .070, .060, and .080 for FI_W on

X, .052, .070, .045, and .075 for FS_W on X, .053, .067, .078, .162, for FI_w on W,

and .061, .072, .092, and .153 for FS_w on W. We observed that at CN=10 the Type I error

was greatly inflated for all the growth factors on X and on W in both True model and

1-MLR models. On the other hand, the Type I error rate for FI_W on X and FS_W on X at

CN=100, 50, and 30 were within the liberal range of robustness. The Type I error rate for

FI_B on W and FS_B on W or FI_w on W and FS_w on W at CN=100 and 50 were also within the

liberal range of robustness but those at CN=30 and 10 were biased and fell out the range of

robustness. Our findings are consistent with Muthén and Satorra (1995) and suggest that

cluster numbers as small as 50 be sufficient to avoid distortion of result due to complex

sampling and to adjust for deviation of data normality in the variables.

4.6 Discussion

The test statistics and fit indices cannot reflect the information for a neglected

higher-level structure under the study design. First, according to the test statistics, the

smaller χ^2 test statistic value gave us a false impression of better model fit when we

neglected the higher level modeling in the dependent data set. In fact, the χ^2 test statistic

values decreased as the number of degrees of freedom decreased. So, the smaller χ^2 test

statistics resulted from the smaller number of degrees of freedom multiplied by the

discrepancy function. For model fit indices, all models had CFI values that exceeded the

conventional cutoff values; and thus, CFI had no discernment of model misspecification.

By mathematical definition, SRMR allows us to know the mean of standardized

differences between elements in the model-driven variance-covariance matrix and the

elements in observed variance-covariance matrix (P. M. Bentler, 1995). All the SRMR

values, including SRMRW, SRMRB, and SRMR, were below the conventional cutoff

scores (SRMR<0.05), suggesting SRMR provides no information of an ignored higher

level, either. Likewise, RMSEA suggested a good fit of the model to the data except that

the True model had the highest RMSEA value at CN=10 and CS=5 (RMSEA = .10). Even

though the True model was the correct model, a bad model fit was obtained for small

sample size (n=50). Our result was congruent with Hox's suggestion (1995) that sample

size should be at least 50, preferably 100, for multilevel models.

In terms of regression coefficients, the parameter estimates of the fixed effects

remained unbiased across the five models. The greater the sample size, the more accurate

and efficient the regression coefficients. Their corresponding standard error, 95% coverage

rate, and empirical power were also congruent to each other, except that the 1-ML-model

underestimated the SEs of the regression coefficients for growth factors on W, gave a

shrunken 95% coverage rate and inflated empirical power at smaller cluster numbers,

while the design-based approaches (i.e. 1-MLR models) produced consistent standard error

estimates as what the model-based approach did (i.e. True model). In addition, by

introducing the cluster-level covariate in the 1-MLR-XW model, we can obtain estimates

of regression coefficients for the within-level growth factors (i.e. FI_W and FS_W) on W,

which are consistent with those for the between-level growth factors (i.e. FI_B and FS_B) on

W with only slightly inflated SEs and almost negligible power loss.

Results for the covariance of growth factors showed the greatest difference among

the three model specifications. The covariance between growth factors indicates the degree

to which the initial status predicts the rate of linear change over time (Kline, 2005). The

covariance estimates of the True model, the 1-MLR-XW, and the 1-ML-XW models were

close to the population value, zero, while those from the 1-MLR-X and 1-ML-X model

were upwardly biased. Besides, the 1-ML models had underestimated SEs for the

covariance estimates, inconsistent (lowered) 95% coverage rate, and inflated Type I error

rate. The biased covariance estimates along with underestimated standard error in

1-MLR-X and 1-ML-X model led to erroneous statistical inference in that high initial

status predicts higher linear increase rate when in fact there was no relationship between

initial status and linear increase. On the contrary, with the integration of the cluster-level

predictor W to the model, the 1-MLR-XW model produced covariance estimates and the

corresponding 95% CI coverage rate and empirical power consistent with those from the

True model. In sum, incorporating predictors from an ignored level in the model along with

adjusted standard errors for the parameter estimates can yield consistent covariance

estimates between growth factors and avoid incorrect statistical inferences.

As for the residual variance of growth factors, the distribution of residual variance

varied among models. The Two-level model produced residual estimates consistent with

the population value, while the 1-MLR models and 1-ML models generated upwardly

biased variance estimates. The biased variance estimates can be explained as the following.

For data of multilevel structure (e.g. a two level structure), the total residual variance of

growth factors should be the sum of the within-level and between-level residual variance of

growth factors. Because we neglected modeling the data dependency using a model-based

approach, the between level residual variance cannot be the estimated separately; as a

result, the 1-ML and the 1-MLR models produce total variance component of residuals,

including the variance components in both levels and the variance components which can

be explained by higher-level covariates. On the contrary, the True model had separate

residual variance estimates from the between- and within-level. For the 1-MLR-XW and

1-ML-XW models, the relationship between covariate W and growth factors were linked;

and thus, the two models estimated a total variance that was approximately the sum of the

residual variance of the between- and within-levels in the True model. In the 1-MLR-X and

1-ML-X models, the relationship between covariate W and growth factors was not established so that the portion of variance unexplained by the covariate W was credited to the residual variance. That was why the 1-ML-X and 1-MLR-X models had residual variance estimates greater than the sum of residual variance of the between- and within-level in the True model. In addition, without standard error correction for the parameter estimates, the 1-ML models underestimated SEs for the factor residual variances.

As for mean structure, compared with the True model and the two 1-MLR models, the 1-ML models underestimated the SEs for the growth factor mean structures, produced inconsistent 95% coverage rate, and had inflated empirical power at small cluster number settings. The 1-MLR-X and 1-MLR-XW models had negligible differences in the mean structure results.

5. CONCLUSIONS AND SUGGESTIONS

The first study compared the similarities and differences of analyzing complex survey data with equal/unequal multilevel structures using a design-based single-level CFA model (the one-level model) and two model-based multilevel CFA models (the two-level true model and the two-level maximum model) on the overall model fit indices, the parameter estimates, 95% coverage for both fixed effects and random effects, and the statistical inferences on detecting the parameter estimates.

Our simulation study showed that the one-level model (the design-based model) provided satisfactory results only under equal between/within structures. However, under the simple between-/ complex within-level structure, the one-level model yielded erroneous cross-loaded factor loadings estimates and biased random effect estimates. As the between-level structure became more complicated than the within-level structure (i.e. the Scenario 3: complex between-/simple within-level structure), the design-based approach produced biased single- and cross-loaded pattern coefficient estimates and poor random effect estimates.

Modeling the data structure as it is (i.e., using the two model-based multilevel

models) turned out to be a better analytical strategy for analyzing multilevel data. However,

the higher level model structure may not always be the focus or interest of a study in which

no specific hypothesized model is set for the higher level. This may also be the reason why

the design-based single level approach, as indicated previously, is a commonly used

approach for analyzing multilevel data given the advantage of simplicity of this approach

(i.e., only one model is needed for specification). Under such circumstance, constructing

multilevel models can be difficult and daunting for researchers with limited higher-level

information from the available data, theories, or prior research. The two-level maximum

model, where the between-level model is a saturated model (i.e. estimating all the unique

non-directional parameters in the between-level model), is a better and feasible alternative

than the design-based one-level approach and even the model-based two-level model with

the requirement for specifying different hypothesized models at different levels.   Thus, if

the researcher's focus is only on validating the pooled-within covariance structure of

complex survey data (in other words, only a within-level model is of interest and

hypothesized), we recommend the use of the two-level maximum model instead of the

one-level model with robust standard estimator (i.e. design-based approach) for multilevel

data because more consistent and efficient model parameter estimates can generally be obtained through the maximum model strategy.

In educational, behavioral, and organizational research, it is common to have a hierarchical data structure, especially in the format of complex survey data and panel data. Information from the higher level is not readily available for some purposes. The first goal of the second study aimed to answer what the effects are if we ignore the highest data-level on the criterion variables. Based on the simulation result, the two 1-MLR models and the two 1-ML models yielded fixed effect parameter estimates consistent with those from the True model. However, in the regular LGCM with a higher level covariate (i.e. 1-ML-XW model), the standard error of the regression weight estimates for the growth factors on higher-level covariate were underestimated, the corresponding 95% coverage rate was shrunk, and the empirical power was inflated, especially at smaller cluster numbers. Our findings were consistent with previous research (Luo & Oi-man Kwok, 2009; Moerbeek, 2004) that when a higher level structure is neglected the standard errors of the regression coefficient from the neglected level are underestimated. However, there is no practical solution to the problem. We suggested that placing cluster-level covariates into the 1-level model along with the use of design-based robust standard error estimator provide estimates

of regression weights and standard errors that were consistent with those from the true

MLGCM.

We also observed a major difference on the growth factors covariance estimates

and residual variances. Therefore, to address our second goal of this paper, we found the

regular LGCMs with or without the cluster-level predictor had underestimated SEs for the

covariance estimates, inconsistent (lowered) 95% coverage rate, and inflated Type I error

rate. For the design-based MLGCM models, including the cluster-level predictor or not

had a great impact on the factors covariance estimates and residual variance estimate. The

cluster-level predictor in the 1-MLR-XW model restored the covariance estimates

distorted by the 1-MLR-X model and accounted for the unexplained variance in the factor

residual variance. Researchers need to exert caution on the biased growth factor covariance

and inflated residual variance estimates as well as their erroneous corresponding statistical

inferences due to ignoring the dependent nature of data when the cluster-level covariate

was not included and the standard error for parameter estimates were not adjusted.

For the sample size issue, we found that when sample size is too small (e.g. n=50)

the MLGCM failed to have a good RMSEA model fit index. The reason was that the

maximum likelihood estimator was a large sample size estimation method and the

dependent data with insufficient sample size resulted in the biased parameter estimates,

especially in the higher-level structures. Owing to this reason, the biased parameter

estimates incurred inaccurate cluster-level model-implied variance-covariance matrix and

distorted total model-implied variance-covariance matrix. The larger disparity between the

observed and model-implied variance-covariance due to the smaller sample size resulted in

the larger value of population discrepancy function per degree of freedom, and the thus

worse RMSEA value. Moreover, small cluster number (e.g. CN=10) also resulted in

reduced power on the growth factors' mean estimates and the regression coefficient

estimates between the cluster-level covariate (e.g. W) and growth factors in the True model

and 1-MLR models. Additionally, a cluster number less than 50 also caused an inflated

Type I error rate for the fixed effect estimates of higher-level covariate. Therefore, our

study was consistent with Muthén and Satorra (1995) to suggest that researchers have a

cluster number at least as large as 50, especially for 3 or 4 waves of repeated measurement

to perform a MLGCM.

As a concluding remark, we encourage researchers to include predictors from the

ignored level if the global cluster-level covariates are available in the complex survey data

and perform a design-based MLGCM. The guideline in choosing the global cluster-level

covariates should be based on that the covariates help explaining the growth factors with the support from the literature, theories, or researchers' experiences. For exploratory research, the statistical significance of the regression coefficient of the higher-level covariate can be used to decide whether the covariates should be included in the model or not. Another overall model selection method, that is, the chi-square difference test between the full model (e.g. the 1-MLR-XW model with the regression coefficients of growth factors on W freely estimated) and restricted model (e.g. the 1-MLR-XW model with the regression coefficients between of growth factors on W set as zero) can be used to determine whether the cluster-level covariate can statistically and practically improve the integrity of the proposed model.

In many studies, the higher-level covariate and the lower-level covariate may or may not correlate with each other. For example, averaged school-level SES is associated with individual-level SES but is not associated with the gender or ethnicity of the individual. One limitation of study two is that the correlation pattern between higher-level covariate W and the lower-level covariate X is not considered as a design factor in the simulation study. Future research can be conducted assuming higher-level and lower-level covariates with different correlation patterns.

REFERENCES

Agrawal, A., & Lynskey, M. T. (2007). Does gender contribute to heterogeneity in criteria for cannabis abuse and dependence? Results from the national epidemiological survey on alcohol and related conditions. *Drug and Alcohol Dependence*, *88*, 300-307.

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, *149*(1), 1-43.

Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological Methodology*, *17*, 71–103.

Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, *18*(3), 1453–1463.

Anderson, D. A., & Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society. Series B (Methodological)*, *47*(2), 203-210.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Arbuckle, J. L. (2003). *Amos 5 user's guide.* Chicago, IL: SPSS.

Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference*.

Au, K., & Cheung, M. W. (2004). Intra-cultural variation and job autonomy in 42 countries. *Organization Studies*, *25*(8), 1339.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, *31*(1), 419–456.

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

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

Blakely, T. A., & Woodward, A. J. (2000). Ecological effects in multi-level studies. *Journal of Epidemiology Community Health*, *54*(5), 367-374.

Bock, R. D. (1960). Components of variance analysis as a structural and discriminal analysis for psychological tests. *British Journal of Statistical Psychology*, *13*, 151–163.

Bock, R. D., & Bargmann, R. E. (1966). Analysis of covariance structures. *Psychometrika*, *31*(4), 507–534.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*, 303-316.

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

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (p. 149). Mahwah, NJ: Lawrence Erlbaum Associates.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 141-152.

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A., Carlo, M., & Francis, D. J. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology*, *98*, 170-181.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83.

Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 193–208.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models for social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage.

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

Cheung, M. W. (2007). Comparison of methods of handling missing time-invariant covariates in latent growth models under the assumption of missing completely at random. *Organizational Research Methods*, *10*(4), 609-634.

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*(2), 347–357.

Cohen, P., Cohen, J., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cronbach, L. J., & Webb, N. (1975). Between-class and within-class effects in a reported aptitude X treatment interaction: Reanalysis of a study by G. L. Anderson. *Journal of Educational Psychology*, *67*, 717-724.

Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, *12*(1), 45–57.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529-569.

Curran, P. J., & Hussong, A. M. (2002). Structural equation modeling of repeated measures data: latent curve analysis. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 59-85). Mahwah, NJ: Lawrence Erlbaum Associates.

Davidov, E., Yang-Hansen, K., Gustafsson, J. E., Schmidt, P., & Bamberg, S. (2006). Does money matter? A theory-driven growth mixture model to explain travel-mode choice with experimental data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *2*(3), 124–134.

De Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, *32*(1), 132-150.

De Leeuw, J., & Kreft, I. G. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics*, *20*(2), 171.

Dickinson, L. M., & Basu, A. (2005). Multilevel modeling and practice-based research. *Annals of Family Medicine*, *3*(1), 52-60.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford, UK: Clarendon Press.

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods. Vol. 1(3)*, *1*(3), 236-248.

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

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, *35*(2), 333–363.

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

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling concepts, issues, and applications*. Quantitative methodology series. Mahwah, NJ: Lawrence Erlbaum Associates.

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*, 149-167.

Enders, C. K. (2008). A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(3), 434-448.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430–457.

Everson, H. T., & Millsap, R. E. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist*, *39*, 157-172.

Fan, X. (1997). Canonical correlation analysis and structural equation modeling: What do they have in common? *Structural Equation Modeling*, *4*(1), 65-79.

Fassinger, R. E. (1987). Use of structural equation modeling in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 425–436.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*(4), 409-420.

Gall, M. D., Gall, J. P., & Borg, W. R. (2006). *Educational research: An introduction* (8th ed.). New York: Prentice Hall.

Goldstein, H. (1987). *Multilevel models in education & social research*. Port Jervis, NY: Lubrecht & Cramer, Limited.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). New York: Wiley & Sons.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*(4), 455-467.

Graham, J. M. (2008). The general linear model as structural equation modeling. *Journal of Educational And Behavioral Statistics*, *33*(4), 485-506.

Graham, J. W. (2003). Adding missing-data-relevant variables to fiml-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 80.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwood, CT: Information Age Publishing.

Hardin, J. W., & Hilbe, J. M. (2002). *Generalized estimating equations* (1st ed.). Boca Raton, Florida: Chapman & Hall.

Hardin, J. W., & Hilbe, J. M. (2007). *Generalized linear models and extensions* (2nd ed.). College Station, TX: Stata Press.

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

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, *31*(2), 423–447.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*(6), 723.

Holt, D., Scott, A. J., & Ewings, P. D. (1980). Chi-squared tests with survey data. *Journal of the Royal Statistical Society. Series A (General)*, *143*(3), 303–320.

Hox, J. J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In *Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 141-156). Nijmegen: ITS.

Hox, J. J. (1995). *Applied multilevel analysis* (2nd ed.). Amsterdam, Netherlands: TT-Publikaties.

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

Hox, J. J., & Kleiboer, A. M. (2007). Retrospective questions or a diary method? A two-level multitrait-multimethod analysis. *Structural Equation Modeling: A*

Multidisciplinary Journal, *14*(2), 311–325.

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

Hox, J. J., & Maas, C. J. M. (2004). Multilevel structural equation models: The limited information approach and the multivariate multilevel approach. In K. V. Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: Theory and applications*. The Netherlands: Kluwer Academic Publishers.

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

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351–362.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 221-233.

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6-23.

Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, *42*(4), 805-820.

Jöreskog, K. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443-482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183-202.

Jöreskog, K. G. (1970a). A general method for analysis of covariance structures. *Biometrika*, *57*(2), 239-251.

Jöreskog, K. G. (1970b). *A general method for estimating a linear structural equation system* (No. RB-70-54) (p. 43). Princeton, NJ: Educational Testing Service.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York, NY: Seminar Press.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In *Applications of statistics*. Amsterdam, Netherland:

North-Holland.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 433-477.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS, Inc.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 user's guide* (1st ed.). Lincolnwood, IL: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide* (2nd ed.). Lincolnwood, IL: Scientific Software.

Kano, Y. (1992). Robust statistics for test-of-independence and related structural models. *Statistics & Probability Letters*, *15*(1), 21–26.

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

Kaplan, D. W. (2008). *Structural equation modeling: foundations and extensions*. Thousand Oaks, CA: Sage Publications.

Kashy, D. A., Donnellan, M. B., Burt, S. A., & McGue, M. (2008). Growth curve models for indistinguishable dyads using multilevel modeling and structural equation modeling: the case of adolescent twins' conflict with their mothers. *Developmental Psychology*, *44*(2), 316-329.

Kendall, M. G., & Stuart, A. (1979). *The advanced theory of statistics.* (Vol. 2). New York, NY: Macmillan.

Khoo, S. T., West, S. G., Wu, W., & Kwok, O. (2006). Longitudinal methods. In M. Eid & E. Dienner (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp. 301-317). Washington, DC: APA.

Kish, L. (1995). *Survey sampling*. Malden, MA: Wiley-Interscience.

Kish, L., & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(1), 1-37.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology*, *86*(1), 3–16.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.

Kreft, I. G. G. (2006). Are multilevel techniques necessary? An overview, including simulation studies. *California State University, June*, *25*, 1996.

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974.

Lee, E. S., & Forthofer, R. N. (2006). *Analyzing complex survey data*. Newbury Park, CA: Sage.

Lee, S. Y. (1986). Estimation for structural equation models with missing data. *Psychometrika*, *51*(1), 93–99.

Lee, S. Y., & Song, X. Y. (2007). A unified maximum likelihood approach for analyzing structural equation models with missing nonstandard data. *Sociological Methods & Research*, *35*(3), 352.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberber, O. (2006). *SAS for mixed models, second edition* (2nd ed.). SAS Publishing.

Llabre, M. M., Spitzer, S., Siegel, S., Saab, P. G., & Schneiderman, N. (2004). Applying latent growth curve modeling to the investigation of individual differences in cardiovascular recovery from stress. *Psychosomatic Medicine*, *66*(1), 29-41.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*(4), 817-827.

Longford, N. T. (1993). *Random coefficient models*. Oxford, UK: Clarendon Press.

Longford, N. T., & Muthén, B. O. (1992). Factor analysis for clustered observations. *Psychometrika*, *57*, 581-597.

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

Luke, D. A. (2007). Multilevel growth curve analysis for quantitative outcomes. In S. Menard (Ed.), *Handbook of longitudinal research* (pp. 545-564). Amsterdam, Netherlands: Elsevier.

Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing

cross-classified data. *Multivariate Behavioral Research*, *44*(2), 182-212. doi:10.1080/00273170902794214

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 86–92.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201-226.

Mathews, C., Aaro, L. E., Flisher, A. J., Mukoma, W., Wubs, A. G., & Schaalma, H. (2009). Predictors of early first sexual intercourse among adolescents in Cape Town, South Africa. *Health Education Research*, *24*(1), 1.

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

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64-82.

Mehta, P. D., & Foorman, B. R. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading*, *9*, 85-116.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259-284.

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, *41*(4), 473–497.

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, *39*(1), 129-149.

Mooijaart, A., & Bentler, P. M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica*, *45*(2), 159–171.

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

Muthén, B. O. (1990). Mean and covariance structure analysis of hierarchical data. *UCLA Statistics Series*, *62*.

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

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*(3), 376-398.

Muthén, B. O., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus*. Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., & Asparouhov, T. (2002). *Using Mplus monte carlo simulations in practice: A note on non-normal missing data in latent variable models* (No. 2). Mplus Web Notes. CA: Muthén & Muthén.

Muthén, B. O., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, *31*(6), 1050–1066.

Muthén, B. O., & Asparouhov, T. (2009). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In *Handbook of Advanced Multilevel Analysis*. New York, NY: Routledge.

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

Muthén, B. O., Khoo, S. T., & Gustafsson, J. E. (1997). *Multilevel latent variable modeling in multiple populations*.

Muthén, B. O., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. *Multilevel analysis of educational data*, 87–99.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267-316.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus V5.21*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus User's Guide. Fifth Edition.* Los Angeles, CA: Muthén & Muthén.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel structural equation modeling. In S. Lee (Ed.), *Handbook of latent variable and related models* (pp. 209-227). Amsterdam, Netherlands: Elsevier.

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

Rigdon, E. (1998). Structural equation modeling. In G. A. Marcoulides (Ed.), *Modern

methods for business research (pp. 251–294). Mahwah, NJ: Lawrence Erlbaum Associates.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351-357.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, *50*(2), 203–228.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology*, 249–278.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *Proceedings of the business and economic statistics section* (pp. 308-313).

Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Comput. Stat. data anal.*, *10*(3), 235-249.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507-514.

Schmidt, W. (1969). *Covariance structure analysis of the multivariate random effects model* (Unpublished doctoral dissertation). University of Chicago.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. Wiley Series in Probability and Statistics (1st ed.). New York, NY: Wiley-Interscience.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*, *86*(2), 420–428.

Skinner, C. J. (1989). Domain means, regression and multi-variate analysis. In C. J. Skinner, D. Holt, & T. M. F. Smith (Eds.), *Analysis of complex surveys* (pp. 165-190). New York, USA: Wiley.

Skinner, C., Holt, D., & Wrigley, N. (1997). *The analysis of complex survey data*. Hoboken, NJ: John Wiley & Sons Inc.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: SAGE Publications.

Stapleton, L. M. (2006). Using multilevel structural equation modeling techniques with complex sample data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing.

Stapleton, L. M. (2008). Analysis of data from complex surveys. In E. D. de Leeuw, J. J.

Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 342-369). New York, NY: Lawrence Erlbaum Associates.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling*, *7*(2), 149–162.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of factors. In *annual spring meeting of the Psychometric Society*. Iowa City, IA.

Stoolmiller, M. (2007). Latent growth curve models. In S. Menard (Ed.), *Handbook of longitudinal research* (pp. 253-544). Amsterdam, Netherlands: Elsevier.

Thompson, B. (2000). Canonical correlation analysis. In *Reading and understanding more multivariate statistics*. Washington, DC: American Psychological Association.

du Toit, S. H., & du Toit, M. (2008). Multilevel structural equation modeling. In *Handbook of multilevel analysis* (pp. 435–78).

Van Landeghem, G., De Fraine, B., & Van Damme, J. (2005). The consequence of ignoring a level of nesting in multilevel analysis: A comment. *Multivariate Behavioral Research*, *40*(4), 423-434.

Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, *5*(4), 425–433.

Watt, H. (2008). A latent growth curve modeling approach using an accelerated longitudinal design: the ontogeny of boys' and girls' talent perceptions and intrinsic values through adolescence. *Educational Research and Evaluation*, *14*(4), 287-304.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817-838.

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

Yuan, K. H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, *40*(1), 115-148.

Yuan, K. H., & Bentler, P. M. (1997). Mean and covariance structure analysis:

Theoretical and practical improvements. *Journal of the American Statistical Association*, 767–774.

Yuan, K. H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *51*(2), 289–310.

Yuan, K. H., & Bentler, P. M. (1999). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions. *Statistica Sinica*, *9*, 831–854.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165-200.

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

Yuan, K. H., & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*, *65*(2), 245–260.

VITA

Name: Jiun-Yu Wu

Address: Dept. of Educational Psychology

                    c/o Dr. Victor Willson

                    Texas A&M University

                    College Station, TX 77843-4225

Email Address: jiunyu.rms@gmail.com

Education: B.S., Communication Engineering, National Chiao Tung University,

                    Taiwan, 2001

                    M.S., Communication Engineering, National Chiao Tung University,

                    Taiwan, 2003

                    Ph.D., Educational Psychology, Texas A&M University, USA, 2010