Measurement Invariance across Groups in Latent Growth Modeling

Eun Sook Kim  
*University of South Florida, ekim3@usf.edu*

Victor L. Willson  
*Texas A & M University - College Station*

Follow this and additional works at: [http://scholarcommons.usf.edu/edq_facpub](http://scholarcommons.usf.edu/edq_facpub)

Part of the [Educational Assessment, Evaluation, and Research Commons](http://scholarcommons.usf.edu/edq_facpub/1)

Scholar Commons Citation


[http://scholarcommons.usf.edu/edq_facpub/1](http://scholarcommons.usf.edu/edq_facpub/1)

This Article is brought to you for free and open access by the Educational Measurement and Research at Scholar Commons. It has been accepted for inclusion in Educational Measurement and Research Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Measurement Invariance Across Groups in Latent Growth Modeling

Eun Sook Kim\textsuperscript{a} & Victor L. Willson\textsuperscript{b}

\textsuperscript{a} University of South Florida
\textsuperscript{b} Texas A&M University

Published online: 09 Jun 2014.


To link to this article: http://dx.doi.org/10.1080/10705511.2014.915374
Measurement Invariance Across Groups in Latent Growth Modeling

Eun Sook Kim\(^1\) and Victor L. Willson\(^2\)

\(^1\)University of South Florida

\(^2\)Texas A&M University

This Monte Carlo study investigated the impacts of measurement noninvariance across groups on major parameter estimates in latent growth modeling when researchers test group differences in initial status and latent growth. The average initial status and latent growth and the group effects on initial status and latent growth were investigated in terms of Type I error and bias. The location and magnitude of noninvariance across groups was related to the location and magnitude of bias and Type I error in the parameter estimates. That is, noninvariance in factor loadings and intercepts was associated with the Type I error inflation and bias in the parameter estimates of the slope factor (or latent growth) and the intercept factor (or initial status), respectively. As noninvariance became large, the degree of Type I error and bias also increased. On the other hand, a correctly specified second-order latent growth model yielded unbiased parameter estimates and correct statistical inferences. Other findings and implications on future studies were discussed.

**Keywords:** factorial invariance, latent growth model, measurement invariance, multiple group analysis, second-order latent growth model

Data are often collected at multiple time points. With multiple time-point data, the changes over time of an outcome variable are usually of focal interest. Especially in social science, including educational and psychological research, longitudinal data are commonly found with an interest in academic growth or in psychological development over time. Questions in longitudinal studies include what the initial status is, how much change occurs on average, and what the variation among individuals around the average initial status and the average change is (Hancock & Lawrence, 2006).

Research interests in changes, growth, or trends over time have called for longitudinal data analysis, and various data analytic techniques have been developed to handle the data measured at different occasions. Analysis of variance (ANOVA), analysis of covariance (ANCOVA), and their multivariate counterparts (i.e., multivariate analysis of variance and multivariate analysis of covariance) are examples of traditional approaches to repeated measures (Hancock & Lawrence, 2006; Hedeker & Gibbons, 2006).

In a traditional repeated measures analysis, we can model and test the trend over time by incorporating linear or quadratic trend contrasts. More recently, hierarchical linear modeling with time nested within subjects using mixed factors, and latent growth modeling (LGM) with structural equations have been used for longitudinal data analysis.

With the popularity of longitudinal data analysis, several issues have drawn researchers’ attention; for example, modeling the error structure of time within a subject. Often the assumption of no correlation between time points in a univariate repeated measures ANOVA is unreasonable and unrealistic. A number of models to structure the error variance across time have been suggested, such as autoregressive and Toeplitz covariances. Measurement invariance over time is another concern of longitudinal data analysis. When we use a measure over time, we question if the same factor structure holds over time. When a test measures the same construct across time, researchers can compare the results of an identical covariance analysis across time, and the changes over time are interpretable (Wu, Liu, Gadermann, & Zumbo, 2010).
A number of studies have focused on measurement invariance issues in longitudinal data. Meredith and Horn (2001), Oort (2001), and Millsap (2010) called attention to the importance of measurement invariance testing across time (see also Vandenberg & Stanley, 2009; Widaman, Ferrer, & Conger, 2010). The implementation of measurement invariance testing in longitudinal data was also demonstrated with item response theory (IRT) and structural equation modeling (Meade, Lautenschlager, & Hecht, 2005; Millsap, 2010; Oort, Visser, & Sprangers, 2005; Raykov & Amemiya, 2008). Recently, the effects of measurement noninvariance over time on the parameter estimates of latent growth models were extensively investigated (Leite, 2005, 2007; Wirth, 2009). These studies observed bias in the parameter estimates of a latent growth model when the model was constructed with the composite scores or factor scores of noninvariant items over time. Interestingly, when linear growth trend was simulated, the latent growth model with noninvariance showed evidence for a quadratic trend (Leite, 2007).

It should be of note that measurement invariance issues could occur in a multitude of components. For example, when research interests include group differences in the initial performance at baseline and in the growth rate over time (e.g., De Fraine, Van Damme, & Onghena, 2007; McMullin, Wirth, & White, 2007; Palardy, 2008), researchers need to establish measurement invariance both across time and across groups (e.g., Flora, Curran, Hussong, & Edwards, 2008). Meredith and Horn (2001) classified measurement invariance into two categories: (a) cross-sectional measurement invariance (invariance between subpopulations or different groups of people), and (b) longitudinal measurement invariance (or temporal invariance, invariance with the same people at different occasions). They also introduced a combined model of cross-sectional and longitudinal measurement invariance. Some researchers were cognizant of measurement invariance problems that might occur among subpopulations in the use of repeated measures and implemented between-group measurement invariance testing in longitudinal data analysis (e.g., Palardy, 2008). Although the importance of measurement invariance over time has long been advocated and researchers in substantive areas have started conducting measurement invariance testing in longitudinal studies, the behaviors of longitudinal data analysis, specifically, LGM in the presence of measurement noninvariance across groups, has not yet been explicitly investigated through Monte Carlo methods.

This study focused on the issues of measurement noninvariance with respect to group membership in longitudinal data. Under the measurement noninvariance across groups, two models of latent growth were considered for investigation: LGM and second-order latent growth modeling (SOLGM). SOLGM allows the explicit test of measurement invariance by embedding the measurement model of each time point at the first-order level and building the latent growth model at the second-order level (see the next section for details). Given the advantage of SOLGM over LGM in testing measurement quality, it is worth comparing the performance of SOLGM with that of LGM, which inherently assumes measurement invariance. Specifically, the impacts of measurement noninvariance across groups on the parameter estimates of LGM and SOLGM were explored. That is, even when the measurement invariance assumption is met over time, if measurement invariance over groups is not established, how does the measurement noninvariance impact the parameter estimates of LGM and SOLGM, respectively? Three research questions were proposed in this study under the research scenario in which there is no group difference in terms of initial performance and growth.

1. When researchers test group differences in average initial status and average growth over time, but measurement invariance across groups is violated, how does the measurement noninvariance influence the estimated group differences?
2. What is the impact on the parameter estimates of average initial status and average growth when measurement noninvariance exists across groups?
3. What is the implication of the magnitude of measurement noninvariance? Does measurement noninvariance always matter regardless of the size of noninvariance? In other words, when small measurement noninvariance was ignored, the impact on the parameter estimates in LGM was studied and compared with the impact of large measurement noninvariance.

Before the in-depth discussion of methods to address these questions, literature on relevant latent growth models and second-order latent growth models for measurement invariance testing is discussed in the following section.

**LITERATURE REVIEW**

**Latent Growth Modeling**

A longitudinal model can be constructed under different data analytic frameworks. Under hierarchical linear modeling and mixed models, a longitudinal model is expressed as multilevel with time nested within subjects. However, in LGM, the initial performance (intercept) and the growth rate (slope) are considered latent factors, and the observed outcome variable at each time point loads on both intercept and slope factors (see Figure 1). The relation of the intercept and slope factors with the observed scores at each time point is modeled with a set of factor loadings:

\[ T_{it} = \gamma_i + \lambda_i K_t + \epsilon_{it}, \]  

(1)
Testing group difference in latent growth modeling.

By introducing a grouping covariate in LGM, we can test group difference in the initial status (or the intercept factor) and the rate of change (or the slope factor). A binary grouping covariate can take different scores depending on a coding scheme (e.g., 0 or 1 with dummy coding; –1 or 1 with contrast coding). The direct effect of a grouping variable on the intercept factor (a in Figure 1) indicates group difference in initial status. The direct effect of a grouping variable on the slope factor (b in Figure 1) indicates group difference in growth. Multiple groups can be modeled using orthogonal contrasts. Statistical significance testing is commonly employed to show the presence of group difference.

Longitudinal Common Factor Model

When the item scores of a measure at each time point are available, researchers can construct a longitudinal common factor (or measurement) model to evaluate the measurement properties of the repeated measures. The longitudinal common factor model (first-order part in Figure 2) portrays the relation of the measured items to a latent construct as

\[ y_{ijt} = \eta_{ijt} + \epsilon_{ijt}, \]

where \( y_{ijt} \) is the observed score for individual \( i \) variable \( j \) at time \( t \), \( \eta_{ijt} \) is the intercept of variable \( j \) at time \( t \), \( \lambda_{ijt} \) is the factor loading of variable \( j \) at time \( t \), \( \eta_{ijt} \) is the latent factor score of individual \( i \) at time \( t \), and \( \epsilon_{ijt} \) is the residual score of individual \( i \) variable \( j \) at time \( t \) (Grimm & Ram, 2009; Meredith, 1964, for a general common factor model).

The mean structure and the implied variance–covariance matrix at time \( t \) are defined as

\[ E(y_i) = \nu_i + \Lambda_i \kappa_i, \]

\[ \Sigma_i = \Lambda_i \Phi_i \Lambda_i' + \Theta_i, \]

where \( E(y_i) \) is the mean vector of observed variables \( y \), \( \nu_i \) is the vector of intercepts, \( \Lambda_i \) is the factor loading matrix of the items on the latent factors, and \( \kappa_i \) is the vector of means of latent factors; \( \Sigma_i \) is the variance–covariance matrix of observed variables, \( \Phi_i \) is the variance–covariance matrix for factors, and \( \Theta_i \) is the variance–covariance matrix for unique factors.

This longitudinal common factor model not only takes into account measurement error in the model, but also allows researchers to investigate the quality of a measure; for example, if the relations between a factor and items (factor loadings) are substantial and if measurement invariance holds over time or over groups (Ferrer, Balluerka, & Widaman, 2008; Hancock, Kuo, & Lawrence, 2001; Widaman et al., 2010).

FIGURE 1 Latent growth model. I = latent intercept; S = latent slope; g = a grouping covariate.

where \( T_i \) is the observed score of an individual \( i \) at time point \( t \), \( \nu_i \) indicates the intercept of a measure at time \( t \), \( K_i \) indicates a set of latent factor scores of individual \( i \) for two factors (i.e., latent intercept and latent slope), \( \lambda_i \) is a set of factor loadings at time \( t \) for two factors, and finally \( \epsilon_{ijt} \) is the error score of individual \( i \) at time point \( t \).

In the linear growth model, each factor loading on the slope factor is fixed at a certain value representing the time interval from the previous occasion. For example, if a measure was administered four times every school year, the factor loadings of time \( t \) can be 0, 1, 2, and 3, reflecting the first through the fourth school year. We can model the factor loadings with different values depending on the time interval. For example, if the data were collected throughout 5 years skipping the third year, one possible set of factor loadings is 0, 1, 3, and 4. The factor loadings can also be freely estimated. The slope factor represents the average rate of change or growth over a designated period of time. The variance of the slope factor signifies the individual variation in growth. On the other hand, all factor loadings of observed outcome variables on the intercept factor are constrained at 1.0. Then, one possible factor loading matrix for the intercept and slope factors is:

\[ \Lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}. \] (2)

The intercept factor captures the initial status on average at the first time point if the corresponding factor loading is fixed at 0. The variance of the intercept factor indicates the variation across individuals in initial status.
Measurement invariance. When a measure is used repeatedly over time or across different groups of people, the same construct is assumed to be measured across time or across groups (Meredith & Horn, 2001). The measurement invariance over time means the equivalence of the terms in the common factor model across time. That is,

\[ \nu_1 = \nu_2 = \ldots = \nu_t = \nu, \]  

(6)

\[ \Lambda_1 = \Lambda_2 = \ldots = \Lambda_t = \Lambda, \]  

(7)

\[ \Theta_1 = \Theta_2 = \ldots = \Theta_t = \Theta. \]  

(8)

Note that a general term for measurement invariance was consistently used in this study, but, precisely speaking, under the common factor model the defined measurement invariance is called factorial invariance (Meredith, 1964). Depending on which terms are invariant across time, a different level of factorial invariance is established (Meredith, 1993; Widaman & Reise, 1997). With the equivalence of factor loadings (\( \Lambda_1 = \Lambda_2 = \ldots = \Lambda_t = \Lambda \)), weak factorial invariance holds over time. Strong factorial invariance over time requires the equality of intercepts (\( \nu_1 = \nu_2 = \ldots = \nu_t = \nu \)) across occasions in addition to the invariant factor loadings. When the error variances are additionally equal over time (\( \Theta_1 = \Theta_2 = \ldots = \Theta_t = \Theta \)), strict factorial invariance is established. For legitimate interpretation of change over time, at least strong factorial invariance of the repeated measures is required.

In addition, we can test the measurement invariance across groups by building the aforementioned common factor model for each group and constraining the terms in the model equal across \( g \) numbers of groups. That is,

\[ \nu_1 = \nu_2 = \ldots = \nu_g = \nu, \]  

(9)

\[ \Lambda_1 = \Lambda_2 = \ldots = \Lambda_g = \Lambda, \]  

(10)

\[ \Theta_1 = \Theta_2 = \ldots = \Theta_g = \Theta. \]  

(11)
Second-Order Latent Growth Modeling

The observed variable at each time point ($T_i$ in Equation 1) in the latent growth model is typically composite scores (i.e., either sum or mean) or factor scores of a set of items in a measure (Leite, 2007; Meredith & Horn, 2001). One caveat of this common practice (i.e., using composite or factor scores of a measure without investigating measurement properties) is that researchers assume measurement invariance of the observed trajectories across time (Flora et al., 2008) or across groups. However, this assumption is not warranted and should be tested (Meredith & Horn, 2001). As described in Equation 1, regular LGM does not allow researchers to test measurement invariance of the indicators across time or across groups (Widaman et al., 2010).

To evaluate the measurement properties, instead of creating composite scores with item scores under a construct, researchers can integrate a common factor model for the construct (i.e., Equation 3) in LGM, which yields a second-order latent growth model (also called a curve-of-factors model; McArdle, 1988; Meredith & Tisak, 1990). Note that the second-order latent growth model requires item scores to build a common factor model at the first-order level. In the second-order latent growth model, a common factor model is constructed for the measure of each time point. Therefore, the indicators of the intercept and slope factors ($T_i$ in Equation 1) are not the observed composite scores of a measure ($T1$~$T4$ in Figure 1) but the latent factor scores underlying the observed items of a measure ($T1$~$T4$ in Figure 2). The advantages of second-order latent growth model include (a) the partition of the error term at each occasion into the unique factor and the error related to deviations from growth, and (b) the capacity to test measurement invariance (Wänström, 2009).

Considering that measurement invariance is assumed across time and across groups in regular LGM and this assumption is not always met in practice, we conducted a Monte Carlo study to investigate the performance of LGM when measurement invariance across groups is violated. Specifically, the impact of measurement noninvariance across groups on the parameter estimates of LGM was examined when group differences in initial performance and growth are the focal interest of a study. Further, the performance of SOLGM, which permits measurement invariance testing with a longitudinal common factor model, was compared to that of LGM under the same scenario of measurement noninvariance.

METHOD

Research Scenario

To test the impact of measurement noninvariance over groups on latent growth analysis, a Monte Carlo study was conducted. The data generation was based on the following research scenario. For a common factor model, a measure of a single factor with six continuous variables was constructed. Two of six items ($Y_{i2}$ and $Y_{i5}$ where $Y_i$ means the $j$th variable at time $t$) were not invariant across two groups of comparison (i.e., reference and focal groups), which yielded about 33% contamination. Either factor loading or intercept was simulated noninvariant across groups. The population parameters for data generation were selected on the basis of the previous studies in both measurement invariance and latent growth models (Kim & Yoon, 2011; Leite, 2005, 2007; Stark, Chernyshenko, & Drasgow, 2006; Wirth, 2009). The factor loadings of six items on a single factor varied between 0.85 and 1.25 around the mean of 1.0 (Wirth, 2009). Thus, the observed items had different relations to the factor, and subsequently the effects of the items on the computation of composite scores were expected to vary depending on the magnitude of factor loadings. The factor loadings of two noninvariant items for the reference group were 1.25 and 1.15, respectively. The intercepts of six items in the measurement model ranged from −.15 to .25 (Kim & Yoon, 2011; Wirth, 2009) with the mean of zero. The population parameters are presented here:

$$\Lambda = \begin{bmatrix} 1.00 & 0.00 \\ 0.25 & 0.36 \\ 0.85 & -0.15 \\ 0.36 \\ 1.15 & 0.15 \\ 0.36 \end{bmatrix} \quad \tau = \begin{bmatrix} 0.00 \\ 0.25 \\ -0.10 \\ 0.36 \\ 0.15 \\ -0.15 \end{bmatrix} \quad \Theta = \begin{bmatrix} 0.36 \\ 0.36 \\ 0.36 \end{bmatrix}.$$

The adopted population parameters are within the range of commonly chosen simulation parameters in the contexts of similar studies (e.g., Cheung & Rensvold, 2002; Woods & Grimm, 2011) and expected to be observed in real research settings.

This measure was repeatedly administered across four equally spaced time points. Four occasions commonly appeared in longitudinal studies of developmental psychology (Khoo, West, Wu, & Kwok, 2006) and was also adopted in previous simulation studies (e.g., Chen, Kwok, Luo, & Willson, 2010). Measurement invariance was established across time. That is, factor loadings, intercepts, and residual variances were equal over four measurement models. Thus, the same parameters presented earlier were used for all four common factor models. With this data set, a latent growth model was constructed. For each time point, the composite scores of the simulated measure were utilized. Composite scores can be computed by taking either the mean or sum of six items. In this study, the mean of the observed scores was used as composite scores. It should be noted that the composite scores were constructed with six items that included two noninvariant items across groups. The simulation of measurement noninvariance is explained later.

For the latent growth model, mean intercept and mean slope were simulated as 0 and 1, respectively (Leite, 2007).
The covariance matrix ($\Phi$) of the intercept and slope factors is as follows:

$$\Phi = \begin{bmatrix} 0.5 & 0.089 \\ 0.089 & 0.1 \end{bmatrix}. $$

The ratio of the intercept factor variance and the slope factor variance equals 5 to 1 as Muthén and Muthén (2002) recommended. The correlation between two latent factors is .4, a corresponding covariance of 0.089 that we adopted from the studies of Leite (2005, 2007). The population parameters are summarized in Figure 2. In latent growth models, a grouping variable can be entered as a time-invariant covariate and loaded on the time and slope factors to test group difference in the mean of initial performance and the mean of growth rates (paths a and b in Figure 1, respectively). To examine the effects of noninvariance in a measure on the group effects in the latent growth model clearly, no group difference was simulated for the latent intercept mean and the latent slope mean, although two observed variables were noninvariant across groups (i.e., no group mean difference under partial strong invariance). That is, both population parameters corresponding to the group effects on the intercept factor (path a in Figure 1) and on the slope factor (path b in Figure 1) were zero.

**Design Factors**

The following design factors were considered in this study.

**Sample size.** Three levels of total sample size were considered as small ($N = 100$), medium ($N = 200$), and large ($N = 1,000$) to explore the differential effects of measurement noninvariance in the latent growth model due to size. Previous research has recommended minimal sample size of 100 for LGM (Fan, 2003; Leite, 2007). The results from the selected conditions are expected to be applicable to intermediate sample size conditions (e.g., 500).

**Design type.** The groups were designed as either balanced or unbalanced. For a balanced design, the total sample size is equally divided into two groups (e.g., 50 per group). The levels of group size in this study (50, 100, and 500) cover the range of group size for adequate power. According to Wänström (2009), the required group size for sufficient power (i.e., .8) to detect group difference in slope factor was about 300 to 500 for a small effect size ($d = 0.2$) and about 50 to 70 for a large effect size ($d = 0.5$). For an unbalanced design, the ratio between two groups was set at 4 to 1. Thus, the samples were simulated with the group sizes of 80/20, 160/40, and 800/200. The large sample size served as the reference group. The small group size (20% of the total sample) was determined to be at the extreme for unbalanced but statistically estimable levels because low convergence and inaccuracy were reported in imbalance proportions of 10% or less (Henson, Reise, & Kim, 2007; Tofighi & Enders, 2008).

**Location of measurement noninvariance.** Measurement noninvariance was simulated either in factor loadings or in intercepts for noninvariant items. This design factor was expected to show the relation between the location of measurement noninvariance and the location of possible bias in the parameter estimates of LGM and SOLGM.

**Number of noninvariant items.** The number of noninvariant items is either zero or two of six items. First, complete invariance across groups was simulated to establish basal Type I error and assumed nonbias under invariance. It is expected that under complete invariance, Type I error will be near the nominal level and the parameter estimates in the model will be within sampling error to population parameters. In the second scenario, two of six items were noninvariant, which yielded 33% contamination.

**Magnitude of noninvariance.** The magnitude of noninvariance varied with three levels: small, large, or counterbalanced. For factor loadings, .3 was subtracted uniformly for both noninvariant items of the focal group, which yielded a large effect size. For small size of noninvariance in factor loadings, .15 was deducted. As to intercepts, .5 for large differential item functioning (DIF) or .25 for small DIF was added uniformly to both item intercepts. These differences between two groups are considered as either large or small in the measurement invariance literature (Kim & Yoon, 2011; Stark et al., 2006). The counterbalance in factor loadings was created by subtracting .3 for one noninvariant item, but adding .3 for the other noninvariant item of the focal group. The counterbalance in intercept was simulated in the same way with the magnitude of .5.

**Fitted Models**

Two longitudinal models were employed and compared in this study: a latent growth model and a second-order latent growth model as shown in Figures 1 and 2, respectively. Supposing researchers’ interest in group differences in the latent intercept and the latent slope, a grouping covariate was employed for both LGM and SOLGM. For LGM, mean composite scores at each time point were used as indicators and loaded on the intercept and slope factors with the predetermined factor loadings presented in Equation 2, which modeled linear growth.

In SOLGM, the second-order part was fitted identically as LGM. For the first-order common factor model, we considered two variations: strong measurement invariance (strong invariance hereafter) assumed model and partial strong measurement invariance (partial strong invariance hereafter) model. In the strong invariance assumed model the factor...
loading of each item ($Y_{ij} - Y_{ij}$ where $Y_{ij}$ means the $j$th variable at time $t$) was constrained equal between two groups as well as across four time points (e.g., for the first item $Y_{i1}$ in Figure 2, $\lambda_{Y_{1g1}} = \lambda_{Y_{2g1}} = \lambda_{Y_{1g2}} = \lambda_{Y_{2g2}}$, and $g1$ and $g2$ indicate group membership). The same constraints were imposed on the intercepts. In the partial strong invariance model, strong factorial invariance across groups was withheld and the two noninvariant variables ($Y_{i2}$ and $Y_{i3}$) were correctly specified as DIF. Of note is that strong invariance across time still held in this model (e.g., $\lambda_{Y_{i1}} = \lambda_{Y_{i2}} = \lambda_{Y_{i3}} = \lambda_{Y_{i4}}$). We allowed a group difference in the intercepts of two noninvariant items by introducing a direct effect of group membership on each DIF variable instead of conducting a multiple group analysis to be consistent with the current LGM with a grouping covariate. This proposed model (namely, multiple indicators, multiple causes model) is able to relax or test the intercept noninvariance only unless including an interaction between a common factor and a grouping covariate (Barendse, Oort, & Garst, 2010; Woods & Grimm, 2011). Given this limitation, we applied partial invariance SOLGM to intercept noninvariance conditions only to examine the performance of a correctly specified model that relaxed two DIF items for the group differences in the intercepts. Considering that the partial strong invariance model is a correctly specified model, we expected that Type I error rates are around the nominal level and bias is close to zero. On the other hand, the strong invariance assumed model is expected to behave similarly to a latent growth model because both models falsely imposed strong invariance across time and across groups when the measurement invariance across groups was violated.

To sum up the simulation conditions, first, only two design factors were applicable to the complete invariance conditions: sample size and design type with 6 (3 × 2) conditions. For the noninvariance conditions, the fully crossed design factors (sample size × design type × location of DIF × magnitude of DIF) yielded 36 conditions (3 × 2 × 2 × 3). Thus, the total number of 42 conditions was investigated for LGM and strong invariance assumed SOLGM. However, the location of DIF was limited to intercepts only for the partial strong invariance SOLGM. Under each condition, 500 replications were simulated, which is consistent with other recent longitudinal simulations (e.g., Kwok, Luo, & West, 2010; Wirth, 2009) and produces small variation in the simulation results. Data generation and analysis were conducted with Mplus (version 5.21; Muthén & Muthén, 2009).

Data Analysis

The major interest of this study was in the overlooked measurement noninvariance effects on the parameter estimates in the latent growth model and the second-order latent growth model. Because the unspecified measurement noninvariance across groups was expected to make impacts on the group effects in the latent growth models, the parameter estimates of group effects on the intercept and slope factors (the paths denoted by a and b in Figure 1 for LGM) were our primary focus. The magnitude of each group effect represents a group difference in the mean of each factor (i.e., mean initial status and mean growth rate). We also explored whether the measurement noninvariance affected the other parameter estimates in the latent growth models, specifically, the intercepts of the initial performance and growth factors (paths c and d in Figure 1 for LGM) after controlling for the group effects. To sum up, the parameter estimates examined in this study were (a) group effect on the intercept factor, (b) group effect on the slope factor, (c) intercept of initial status (or latent intercept), and (d) intercept of growth rate (latent growth).

Because researchers generally report statistical significance of the group differences in the latent intercept and slope factors ($p$ values of the paths, a and b, in Figure 1), the statistical significance of each path coefficient at an alpha level of .05 was examined in this study. The default in Mplus is the ratio of a path coefficient to its standard error (namely, $z$ statistic; Est./S.E. in the Mplus output). The corresponding two-tailed $p$ value was referred for statistical significance in this study. Because there was no group difference in the mean intercept and the mean growth, both effects should be statistically nonsignificant. Any statistical significance of these two effects will indicate bias due to the measurement noninvariance across groups. The proportion of cases in which the group difference was statistically significant was calculated for each condition and defined as Type I error.

In addition to the statistical significance of the parameter estimates of the group effects on the mean intercept and the mean growth, the actual deviation from the population parameter was computed as a raw bias. Because statistical significance greatly depends on sample size, the examination of raw bias could provide insights about the actual impacts of measurement noninvariance irrespective of sample size. The raw bias of each simulation condition, $B(\hat{\theta}_c)$ was calculated as:

$$B(\hat{\theta}_c) = R^{-1} \sum_{r=1}^{R} (\hat{\theta}_{rc} - \hat{\theta}_c), \tag{12}$$

where $\hat{\theta}_{rc}$ is the parameter estimate for replication $r$ in condition c, $\theta_c$ indicates the population parameter for $\theta$ in condition c, and $R$ denotes the total number of replications. In this study, $R$ equals 500. When $\theta_c$ equals zero, the raw bias equals the mean of parameter estimates over 500 replications. The raw bias was compared in magnitude across simulation conditions. In addition, the raw bias was interpreted in terms of parameter standard errors. The standard error of each population parameter can be calculated from 500 replications. Thus, standardized bias $SB(\hat{\theta}_c)$ was defined as:
$SB(\theta_c) = \frac{B(\theta_c)}{SE_{\theta_c}},$  \hspace{1cm}(13)$

where $SE_{\theta_c}$ is the standard error of $\theta_c$ (Collins, Schafer, & Kam, 2001; Merkle, 2011). According to Collins et al. (2001), standardized bias over .4 or .5 is expected to have an adverse impact on efficiency and error rates. In this study, standardized bias over .4 was flagged as the presence of bias. In summary, for the aforementioned four parameter estimates, raw bias and standardized bias were computed and interpreted. For group effect parameter estimates, Type I error was also included in the analysis.

RESULTS

The results of this Monte Carlo study are presented in the following order. First, inadmissible solution rates are presented across all simulation conditions. Then, Type I error and bias (raw and standardized) of LGM and strong invariance assumed SOLGM are summarized according to the type of noninvariance: complete invariance, factor loading noninvariance, intercept noninvariance, and counterbalanced noninvariance. The results of LGM and strong invariance assumed SOLGM are reported together because the results showed similar patterns, whereas the performance of partial strong invariance SOLGM is summarized separately at the end.

Inadmissible Solutions

For LGM, in the conditions with at least medium sample size (i.e., 200) all replications converged without computation errors or inadmissible solutions. However, for the small sample size (100), regardless of design type (balanced or unbalanced) about 2% (1.4%–2.2%) of replications showed inadmissible solutions. The source of inadmissible solutions was, by and large, negative variance for the parameter estimates. However, the size of negative variance was very small and could be considered as zero in most cases. With the item scores across time, SOLGM converged without any inadmissible solutions across all simulation conditions including small sample size. In general, LGM and SOLGM under conditions similar to those of this study are expected to yield stable estimates even with a sample size as small as 100.

Type I Error of LGM and Strong Invariance Assumed SOLGM

**Complete invariance across groups.** The Type I error rates are presented in Tables 1 and 2 for zero group difference in the population. For the full invariance conditions, the Type I error rates were around .05 with the range of .032 to .072 for LGM (Table 1) and .036 to .074 for SOLGM (Table 2) regardless of sample size and design type. The observed Type I error rates under invariance appeared to be reasonable because $\alpha$ was .05 in the statistical significance testing. In addition, the ranges of Type I error were within the interval of [.025, .075] computed with an equation, $\alpha \pm 1/2\sigma_\nu$ (Bradley, 1978) and also within 1.96 standard errors of the binomial distribution for $p = .05$ given the sample size of replications for each condition. In some conditions, Type I error was over .05, but still below the maximum of the expected Type I error rates. Overall, in all simulation conditions of complete invariance (i.e., irrespective of sample size and design type), Type I error was likely due to chance. In sum, LGM and SOLGM correctly retained the null hypothesis of no group difference in the mean initial status and the mean growth when measurement invariance across groups was achieved.

**Factor loading noninvariance across groups.** For factor loading noninvariance conditions, Type I error was considerably inflated when the group effects on the latent slope were tested for statistical significance (see the column $S$ on $G$ in Tables 1 and 2). In other words, when there was no group difference in latent growth or change over time, the group difference in latent growth was falsely detected above the chance rate. Type I error inflation became more serious as the magnitude of noninvariance and sample size increased. Figure 3 shows the increasing pattern of Type I error rates in relation to the size of sample and DIF for LGM. In the combination of large sample size (i.e., 1,000) and large noninvariance, the group effects on the latent growth were statistically significant almost all the time (99.4% for balanced design, 97.0% for unbalanced design of LGM; 99.0% for balanced design, 98.40% for unbalanced design of SOLGM). With the large sample size, even small magnitude noninvariance in factor loadings had a substantial impact on the group difference in the mean growth: The Type I error rates for misidentification of a group difference in the mean growth were 66.8% and 52.8% for balanced and unbalanced conditions of LGM, respectively. When the DIF was small in magnitude the Type I error inflation was less salient than with large DIF, shown in Figure 3, in which the graph line of Type I error rate for the small DIF is below the line for the large DIF across all sample sizes. However, in most conditions of LGM and SOLGM, the influence of small noninvariance in factor loadings was not negligible with over 10% Type I error rates.

The Type I error inflation was more serious with balanced conditions. Comparing the Type I error rates between balanced and unbalanced conditions of factor loading noninvariance (see graphs a and b side by side in Figure 3), we observed slightly higher Type I error rates for balanced conditions. In the case of LGM, for balanced conditions with the large DIF, the Type I error rates in the group effects on latent growth ($S$ on $G$ in Table 1) ranged from 37% to
99% depending on sample size, whereas for the unbalanced design, the Type I error rates were between 20% and 97%.

When the factor loadings were not invariant across groups, the statistical significance testing of group effects on the latent intercept exhibited Type I errors about .05 across simulation conditions (see the column I on G in Tables 1 and 2). The range of Type I error rates was [.036, .068] for LGM and [.032, .072] for SOLGM, which are not different from those of full invariance conditions. As shown in Tables 1 and 2, Type I error rates did not vary to any meaningful degree with the different levels of design factors: DIF (either small or large), sample size (100, 200, or 1,000), and design type (balance or unbalanced). The factor loading noninvariance across groups appeared not to make any impact on the statistical significance of group difference in the initial status (or latent intercept).

Overall, LGM and SOLGM showed almost identical patterns of Type I error across simulation conditions. However, compared to LGM, the Type I error rates of SOLGM were slightly higher across conditions.

<table>
<thead>
<tr>
<th>Measurement Location</th>
<th>Noninvariance Magnitude</th>
<th>Sample Size (Group1/Group2)</th>
<th>Type I Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>S on G</td>
</tr>
<tr>
<td>Complete invariance</td>
<td>Zero</td>
<td>50/50</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.042</td>
</tr>
<tr>
<td>Factor loading</td>
<td>Small</td>
<td>50/50</td>
<td>.128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.192</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.668</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.142</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.528</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>50/50</td>
<td>.366</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.572</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.422</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.970</td>
</tr>
<tr>
<td>Counterbalanced</td>
<td></td>
<td>50/50</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.060</td>
</tr>
<tr>
<td>Intercept</td>
<td>Small</td>
<td>50/50</td>
<td>.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>50/50</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td>Counterbalanced</td>
<td>50/50</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100/100</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500/500</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80/20</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160/40</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800/200</td>
<td>.052</td>
</tr>
</tbody>
</table>

Note. S on G = group effect on latent slope; I on G = group effect on latent intercept. Type I error over .100 is shown in bold.
Intercept noninvariance across groups. The intercept noninvariance conditions showed results opposite to those reported earlier (see Tables 1 and 2 and Figure 3) for both LGM and strong invariance assumed SOLGM. In other words, the group effects on the latent intercept showed highly inflated Type I error rates, whereas the Type I error rates of the group effects on the latent slope were near the nominal level (i.e., .05).

The group effect on the mean growth appeared not to be affected by the intercept noninvariance, showing Type I error around .05 with a range of .03 to .07. This range of Type I error rates was consistent with the basal Type I error we observed under the complete invariance conditions and can be interpreted as sampling error. Simulation conditions of sample size, design type, and DIF magnitude were not associated with the Type I error rate variation.

When the group difference in the latent intercept (initial status) was tested for statistical significance under the noninvariance in the intercepts across groups, the Type I error rates were between 19% and 94% for large DIF.
balanced conditions of LGM (24% and 98% for SOLGM) and 11% and 45% for small DIF balanced conditions of LGM (13% and 53% for SOLGM). As observed in the factor loading noninvariance conditions, the Type I error inflation depended greatly on sample size and DIF size. Overall, unbalanced cases showed slightly lower Type I error rates. Even when the size of DIF was small, the group difference in the initial status was often falsely detected, especially when sample size was large (e.g., Type I error of .45 with the sample size 500/500; .28 with 800/200 in the case of LGM). Again the Type I error rates of SOLMG were slightly higher than those of LGM.

Counterbalanced noninvariance across groups. When two noninvariant items were counterbalanced (i.e., one in positive and another in negative directions in the same magnitude), the bias and the Type I error inflation observed in the uniform noninvariance conditions were not detected and the effects of noninvariance were simply null across all simulation conditions. In other words, Type I error was around .05 regardless of sample size and the location of DIF. This applied for both LGM and SOLMG.

Raw and Standardized Bias of LGM and Strong Invariance Assumed SOLGM

Complete invariance across groups. In addition to Type I error, this study examined raw bias and standardized bias (Tables 3 and 4 for LGM). The bias of SOLGM is summarized in text because the bias patterns of SOLGM were almost identical to those of LGM. The tables are available by request. The raw bias was simply zero (.000) or near zero (e.g., 004) under the complete invariance conditions. This is true for all four parameter estimates of LGM and strong invariance assumed SOLGM: the group effects on the latent intercept and growth (I on G and S on G) and the intercepts of the latent intercept and growth (I and S). In terms of standardized bias, not a condition was flagged as showing substantial bias. The standardized bias across all complete invariance conditions was close to zero not exceeding the cutoff, .40. Consistent with Type I error, the parameter estimates in LGM and SOLGM appear to be unbiased across all conditions when measurement invariance holds.

Factor loading noninvariance across groups. Under the factor loading noninvariance conditions, the raw bias was prominent in the parameter estimates of slope: the group effect on latent slope (S on G in Table 3) and the intercept of latent slope (S in Table 3). With respect to these biased parameter estimates, two noticeable patterns emerged. First, the magnitude of raw bias was independent of sample size. Whereas the Type I error rates depended mostly on sample size, the magnitude of raw bias was almost identical across sample size regardless of whether the design was balanced or not. For example, when the DIF was large, the raw bias of the group effects on the latent slope in LGM was about –.10 regardless of sample size and design type.

Second, measurement noninvariance in factor loadings had an impact on both the group effect on the latent slope and the intercept of the latent slope. The raw bias measures of these two parameters were almost identical in size, but in the opposite direction. That is, the intercept of the slope factor was slightly positively biased around .10 on average for the large noninvariance conditions. The raw bias of the group difference in latent growth was estimated at about –.10. Because two factor loadings of the focal group were
simulated to be lower than those of the reference group by .3 for large noninvariance and by .15 for small noninvariance, the negative bias in the group difference in latent growth appears to be reasonable. When DIF was small in factor loadings across groups, the size of bias was approximately half of the bias size of the large DIF conditions (about 0.05). As observed in the large DIF conditions, the group effect on latent growth was negatively biased, whereas the intercept of latent growth was positively biased with the similar magnitude.

Although the raw bias provided useful information about the impact of measurement noninvariance on the parameter estimates, it is challenging to interpret the magnitude of raw bias, and thus, standardized bias was examined (Table 4). Following Collins and colleagues (2001), we set .4 as the cutoff for meaningful standardized bias. As shown in Table 4, with the noninvariance in factor loadings, meaningful standardized bias was observed for the parameters related to the latent slope (i.e., S on G and S). Even with small noninvariance and a small sample size, the parameter
estimates of group effects on the growth factor and of the growth factor intercept were contaminated, and the size of standardized bias was not negligible (\(-0.807\) for \(S\) on \(G\), 0.548 for \(S\) when the design was balanced in LGM; \(-0.910\) for \(S\) on \(G\), 0.865 for \(S\) in strong invariance assumed SOLGM).

On the other hand, the parameter estimates of the group effect on initial performance and the intercept of initial performance appeared to be unbiased with measurement noninvariance of factor loadings of the observed items. The standardized bias as well as the raw bias did not distinguishably deviate from those of the complete invariance conditions. The bias measures were close to zero, indicating that factor loading noninvariance did not have a noticeable influence on the parameter estimates related to the intercept factor.

**Intercept noninvariance across groups.** When the intercepts of the common factor model were noninvariant between the two groups, the group effects on the latent intercept (I on G in Table 3) and the intercept of the latent intercept (I in Table 3) were both biased, with a magnitude
of about .16 for large DIF and about .08 for small DIF for LGM. As observed in the factor loading noninvariance conditions, the sign of bias was the opposite for I on G and for I. The positive group difference in the initial status (I on G) in favor of the focal group was observed, reflecting the simulation scenario in which the intercepts of two focal-group items were created by adding .5 for large DIF to the corresponding intercepts of the reference group. Conversely, the intercept of the intercept factor (I in Table 3) showed negative bias with a similar magnitude of I on G. The group effect on the latent slope and the intercept of the latent slope appeared to be unbiased with near-zero bias across conditions. About raw and standardized bias, the findings of intercept noninvariance conditions were not different from the factor loading noninvariance findings presented earlier except slightly higher raw bias. Hence, the results of raw and standardized bias reported under factor loading noninvariance (e.g., no association with sample size and the direction of bias) can be applied to the intercept noninvariance.

Counterbalanced noninvariance across groups. As observed in Type I error, the counterbalanced noninvariance did not produce bias in the parameter estimates of LGM and SOLGM, although the noninvariance was large. Across all relevant conditions, both raw bias and standardized bias were close to zero, indicating that there was no effect of noninvariance in sum although two items were noninvariant individually.

Partial Strong Invariance SOLGM

Partial strong invariance SOLGM is a correctly specified model in which two noninvariant variables were freely estimated for group difference. When measurement noninvariance was taken into account and group differences were allowed for the noninvariant items by introducing a direct path from the grouping covariate to each DIF item, Type I error was around .05 and both raw and standardized bias were near zero irrespective of simulation conditions (see Table 5). The Type I error inflation and considerable bias of observed in LGM and strong invariance assumed SOLGM were not of concern in partial strong invariance SOLGM.

DISCUSSION

One of the most salient findings of this study is the relationship between the location of the measurement noninvariance and the location of bias and Type I error when measurement invariance was incorrectly imposed in the LGM and strong invariance assumed SOLGM. When noninvariance occurred in the factor loadings of the measure, the parameters related to slope in the latent growth model showed considerable bias and substantial Type I error. On the other hand, noninvariance of the intercept of the measure had an impact on the parameter estimates related to the latent intercept. To put this in another way, the parameter estimates not directly related to the source of noninvariance remained unbiased across all simulation conditions (e.g., the group effect on the slope factor with the intercept noninvariance). For the unbiased estimates, Type I error did not severely deviate from the nominal significance level (.05) irrespective of whether DIF size was small or large; sample size was small, medium, or large; and design was balanced or unbalanced. To sum up, the location of noninvariance matters in LGM and SOLGM. Therefore, identifying where the noninvariance occurs through measurement invariance testing will assist researchers in understanding potential bias in the parameter estimates in LGM.

However, if the noninvariance is counterbalanced with the same size in the opposite direction at the same location, the overall DIF effects at the test level become null as observed in this simulation study because in this case differential test functioning equals the sum of DIF (Roju, van der Linden, & Fleer, 1995). With counterbalanced noninvariance, researchers might not encounter sizable bias or Type I error in the parameter estimates, although substantial noninvariance could exist at the item level. The counterbalanced or compensating noninvariance was discussed in the IRT literature (Roju et al., 1995; Shealy & Stout, 1993; Stark, Chernyshenko, & Drasgow, 2004) and the practice of deleting a biased item without estimating the impact on the overall test bias was cautioned because the item deletion could result in more serious bias testwise.

Considering that group difference is often determined with statistical significance testing in practice, the substantial Type I error rates are of concern. Even when there is no group difference at the initial status and latent growth, researchers might erroneously conclude that there is a group difference due to items with measurement noninvariance. Of even more importance, Type I error rates are not ignorable even with small noninvariance, especially when sample size is large. This indicates the importance of measurement invariance testing and the establishment of measurement invariance across groups in the use of LGM, especially when group differences in latent growth and initial status are the focal interest of study.

The magnitude of Type I error depends on sample size and DIF size: the bigger, the larger (Figure 3). The positive relations between Type I error rates and some design factors (i.e., sample size and DIF size) are not surprising and are consistent with previous studies about measurement invariance under various research settings (Finch, 2005; Kim & Yoon, 2011; Oort, 1998; Stark et al., 2006). The issue of sample size in LGM when noninvariance is present requires more caution in relation to power. For example, Wänström (2009) suggested group size over 400 to detect a small effect size ($d = .20$) between two groups in latent growth. Nonetheless,
in this study a balanced group size of 500 showed near 100% Type I error when factor loadings were not invariant across groups. Therefore, researchers working with large sample data should be more cognizant of the possibility that the observed group effects could be statistical artifacts rather than true differences between groups if measurement invariance is violated.

Unlike Type I error and standardized bias, raw bias was almost constant and independent of sample size and design type. As long as the estimation is stable, the parameter estimates are reasonably not related to sample size and group size. That is, the same model should yield the same or very similar values for parameter estimates regardless of sample size or group size because the same population parameters were used for data generation of different conditions of sample size and group size. However, in statistical analyses, sample size and group size played a role. As sample size and group size increased, both Type I error and standardized bias became more serious. The only design factor related to the magnitude of raw bias among the investigated design factors was the degree of measurement noninvariance. The magnitude of raw bias directly reflected the magnitude of measurement noninvariance either in intercepts or in factor loadings. For instance, when the size of noninvariance doubled from small to large (e.g., from .25 to .50 in the intercepts), the size of raw bias also increased by almost 100% (e.g., from .08 to .16 in I on G, Table 3).

In addition to the magnitude of bias, the direction of bias appeared to be associated with the direction of noninvariance. When noninvariance was simulated in a negative direction in adverse to the focal group (e.g., lower factor loadings for the focal group), the direction of bias was also negative in the estimation of group effects (e.g., S on G in factor loading noninvariance conditions in Table 3). Conversely, the intercept of the latent factor (e.g., the slope factor, S in Table 3) was estimated in a positive direction in favor of the reference group, which, in turn, leads to the negative group effect we observed in S on G. On the other hand, the positive noninvariance in favor of the focal group (e.g., higher intercepts for the focal group) resulted in a positive raw bias (e.g., I on G in intercept noninvariance conditions in Table 3). Instead, the intercept of the corresponding latent factor (e.g., the intercept factor, I in the intercept noninvariance conditions in Table 3) was negatively biased, which plausibly resulted in a positive group effect.

Comparing balanced and unbalanced conditions, Type I error inflation was more serious in the balanced design irrespective of direction (i.e., in favor of or against a certain group). This finding is reasonable because the number of dissimilar cases is maximized when the design is balanced. That is, the effect of noninvariance is likely capitalized when two groups are equally influential. Conversely, when one group is dominant in size, the difference becomes less salient. In other words, larger variability of a covariate (i.e., group membership) in a balanced design than in an unbalanced design leads to smaller standard error for the group effect estimate, which in turn increases the chance of Type I error in this study.

### Table 5

<table>
<thead>
<tr>
<th>Sample Size (Group1/Group2)</th>
<th>S on G</th>
<th>I on G</th>
<th>S</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50/50</td>
<td>.060</td>
<td>.052</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>100/100</td>
<td>.046</td>
<td>.060</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>500/500</td>
<td>.068</td>
<td>.048</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>80/20</td>
<td>.028</td>
<td>.062</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>160/40</td>
<td>.050</td>
<td>.054</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>800/200</td>
<td>.062</td>
<td>.062</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Raw bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50/50</td>
<td>−.003</td>
<td>−.013</td>
<td>.006</td>
<td>.025</td>
</tr>
<tr>
<td>100/100</td>
<td>−.001</td>
<td>.008</td>
<td>.002</td>
<td>−.010</td>
</tr>
<tr>
<td>500/500</td>
<td>.000</td>
<td>−.003</td>
<td>.001</td>
<td>.005</td>
</tr>
<tr>
<td>80/20</td>
<td>.001</td>
<td>−.005</td>
<td>.000</td>
<td>.010</td>
</tr>
<tr>
<td>160/40</td>
<td>−.002</td>
<td>−.001</td>
<td>.003</td>
<td>.003</td>
</tr>
<tr>
<td>800/200</td>
<td>−.002</td>
<td>−.005</td>
<td>.003</td>
<td>.007</td>
</tr>
<tr>
<td>Standardized bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50/50</td>
<td>−.045</td>
<td>−.088</td>
<td>.056</td>
<td>.107</td>
</tr>
<tr>
<td>100/100</td>
<td>−.022</td>
<td>.075</td>
<td>.025</td>
<td>−.061</td>
</tr>
<tr>
<td>500/500</td>
<td>−.086</td>
<td>−.064</td>
<td>.028</td>
<td>.068</td>
</tr>
<tr>
<td>80/20</td>
<td>.013</td>
<td>−.027</td>
<td>.000</td>
<td>.044</td>
</tr>
<tr>
<td>160/40</td>
<td>−.036</td>
<td>−.007</td>
<td>.039</td>
<td>.018</td>
</tr>
<tr>
<td>800/200</td>
<td>−.071</td>
<td>−.082</td>
<td>.086</td>
<td>.093</td>
</tr>
</tbody>
</table>

*Note. S on G = group effect on latent slope; I on G = group effect on latent intercept; S = intercept of latent slope; I = intercept of latent intercept. For simplicity, small and counterbalanced noninvariance conditions are not included, but the results are similar to those of large noninvariance conditions showing around .05 Type I error and near zero raw and standardized bias.*
Finally, the performance of partial strong invariance SOLGM has important implications for longitudinal studies. That is, when measurement noninvariance is correctly specified as DIF in the model, the parameter estimates are unbiased and Type I error is likely to occur by chance. This encourages researchers to test measurement invariance across groups as well as across time in conducting a longitudinal data analysis and to resort to SOLGM, which integrates a longitudinal common factor model and allows modeling partial strong measurement invariance.

IMPLICATIONS FOR FUTURE RESEARCH

Observing substantial bias and Type I error in LGM with noninvariance across groups, we cannot overemphasize the importance of measurement invariance testing. Understanding the issues of measurement noninvariance over groups in LGM raises a couple of questions that call for further research. First, how well is noninvariance across groups detected through measurement invariance testing? With the known problems of measurement noninvariance over groups in LGM, the users of LGM are strongly recommended to conduct measurement invariance testing and to establish measurement invariance across groups. In fact, it is not common, but not difficult, to find a study in educational and psychological research conducting measurement invariance testing over groups with longitudinal data (e.g., Glanville & Wildhagen, 2007; Guttmannova, Szanyi, & Cali, 2008). Once DIF is correctly identified through measurement invariance testing, SOLGM has a capacity to model the DIF between groups (i.e., to specify the noninvariant variables properly), which is expected to yield unbiased estimates and correct statistical inference about the group differences in latent growth and initial performance as demonstrated in this study. However, whether currently available measurement invariance testing methods are able to detect noninvariance across groups properly or not is not thoroughly investigated yet. More specifically, the suitability of the second-order latent growth model as a measurement invariance testing technique needs further examination, especially for the researchers who implement LGM for longitudinal data.

Second, what is the impact of measurement noninvariance across groups on LGM in the presence of measurement noninvariance over time? When between-group noninvariance interacts with temporal noninvariance, how does it affect the performance of longitudinal data analytic methods including LGM? This study assumed measurement invariance over time and investigated the effects of measurement noninvariance across groups only. However, as pointed out earlier, measurement invariance could occur in both ways. Possibly, the patterns of noninvariance over time could be different depending on group membership, which might complicate the effects of measurement noninvariance in the use of longitudinal data. Further research on this topic hopefully will shed a light on the full picture of the relations between LGM and measurement invariance.

CONCLUSIONS

As one of the popular longitudinal data analytic methods, SOLGM has advantages in analyzing individual changes and overall trends over time. One of the advantages of SOLGM is to permit measurement invariance testing by including a common factor model in data analysis. This study found that the consequences of the presence of noninvariant variables across groups could be critical when researchers ignore measurement noninvariance across groups and conduct LGM to test group differences in initial status and latent growth. That is to say, major parameter estimates of LGM with a grouping covariate could be biased depending on where the noninvariance exists. The impacts of factor loading noninvariance on the slope factor and intercept noninvariance on the intercept factor were observed. In addition, Type I error was considerable in statistical significance testing of group difference in initial status and latent growth. The findings of this study hopefully will encourage researchers to conduct SOLGM and ensure measurement quality in the use of longitudinal data.

REFERENCES


