Analyzing individual growth with clustered longitudinal data: A comparison between model-based and design-based multilevel approaches

Hsien-Yuan Hsu¹ · John J. H. Lin² · Susan T. Skidmore³

Published online: 20 June 2017
© Psychonomic Society, Inc. 2017

Abstract To prevent biased estimates of intraindividual growth and interindividual variability when working with clustered longitudinal data (e.g., repeated measures nested within students; students nested within schools), individual dependency should be considered. A Monte Carlo study was conducted to examine to what extent two model-based approaches (multilevel latent growth curve model – MLGCM, and maximum model – MM) and one design-based approach (design-based latent growth curve model – D-LGCM) could produce unbiased and efficient parameter estimates of intraindividual growth and interindividual variability given clustered longitudinal data. The solutions of a single-level latent growth curve model (SLGCM) were also provided to demonstrate the consequences of ignoring individual dependency. Design factors considered in the present simulation study were as follows: number of clusters (NC = 10, 30, 50, 100, 150, 200, and 500) and cluster size (CS = 5, 10, and 20). According to our results, when intraindividual growth is of interest, researchers are free to implement MLGCM, MM, or D-LGCM. With regard to interindividual variability, MLGCM and MM were capable of producing accurate parameter estimates and *SEs*. However, when D-LGCM and SLGCM were applied, parameter estimates of interindividual

variability were not comprised exclusively of the variability in individual (e.g., students) growth but instead were the combined variability of individual and cluster (e.g., school) growth, which cannot be interpreted. The take-home message is that D-LGCM does not qualify as an alternative approach to analyzing clustered longitudinal data if interindividual variability is of interest.

Keywords Clustered longitudinal data · Design-based approach · Model-based approach · Multilevel latent growth curve model

Longitudinal research designs allow for the opportunity to investigate individual development in education and psychology. Particularly, measuring the same individual repeatedly permits researchers to describe how and when attributes of the individual change over time (i.e., intraindividual growth) and whether different individuals change in different ways (i.e., interindividual variability; Curran, Obeidat, & Losardo, 2010; Duncan, Duncan, & Strycker, 2006; Grimm, Ram, & Estabrook, 2016). This description of intraindividual growth outlines the overall individual growth pattern (e.g., growth in mathematics achievement), which is the mean of the trajectory pooling of all the individuals within the sample (Curran et al., 2010). On the other hand, interindividual variability reveals the dissimilarity of the individual growth trajectory (e.g., students differ in their growth in mathematics achievement). Accurate estimation of interindividual variability is extremely important because it provides empirical support for the need to further study explanatory determinants (Duncan et al., 2006; Schaie, 1983; Singer & Willett, 2003) and elicits research questions regarding the individual-related factors accounting for such variability (e.g., what is the role of student-perceived mathematical self-efficacy in the growth of mathematics achievement; Bandura, 1993). Statistically speaking,

✉ Hsien-Yuan Hsu
Hsien-Yuan.Hsu@uth.tmc.edu

¹ Children's Learning Institute, University of Texas Health Science Center at Houston, 7000 Fannin St., Suite 2373I, Houston, TX 77030, USA

² Office of Institutional Research, National Central University, Taoyuan, Taiwan

³ Department of Educational Leadership, Sam Houston State University, Huntsville, TX, USA

obtaining accurate estimates of intraindividual growth and inter-individual variability are essential to understanding development, and thus should be a priority topic for researchers in developmental psychology. Although several studies have been devoted to this topic by examining some influential factors such as sample size, number of time points, and measurement error (e.g., Diallo, Morin, & Parker, 2014; Hertzog, Oertzen, & Ghisletta, 2008; B. O. Muthén & Curran, 1997; Zhang & Wang, 2009), we found little attention has been paid to the methodological issue of how to obtain more accurate estimates of intraindividual growth and interindividual variability within the context of longitudinal data in a hierarchical setting, especially with large-scale studies. Yet there are certainly multiple large-scale data sets available using such a sampling structure. For example, the Longitudinal Study of American Youth (LSAY; Miller, Kimmel, Hoffer, & Nelson, 2000), a national panel study of mathematics and science education in US public schools, adopted a two-stage stratified probability sampling approach – representative schools were randomly selected and students within the selected school were then randomly sampled. This sampling approach produced *clustered longitudinal data* at three levels: cluster (e.g., school), individual (e.g., student), and repeated measures (e.g., number of time points).

Conceptually, individuals within the same cluster are more likely to be more similar to each other (Hox, 2010; Snijders & Bosker, 2012), and, analytically, such individual dependency should be considered to prevent biased estimates of intraindividual growth and interindividual variability (B. O. Muthén, 1997). The consequences of ignoring individual dependency can lead to seriously biased parameter estimates and standard errors (*SEs*), which has been widely discussed within the context of various multilevel models including measurement models (e.g., Julian, 2001; B. O. Muthén & Satorra, 1995; Pornprasertmanit, Lee, & Preacher, 2014; Wu & Kwok, 2012), regression models (e.g., Lai & Kwok, 2015; Moerbeek, 2004; B. O. Muthén & Satorra, 1995), cross-classified models (e.g., Luo & Kwok, 2009; Meyers & Beretvas, 2006), growth mixture models (Chen, Kwok, Luo, & Willson, 2010), and latent growth curve models (e.g., B. O. Muthén, 1997; Wu, Kwok, & Willson, 2015). Although researchers have been well informed by the aforementioned studies, little is known about whether current commonly-used model-based and design-based analytical approaches, which can accommodate individual dependency in data, can produce unbiased parameter estimates of intraindividual growth and interindividual variability given clustered longitudinal data. The current study aimed to close this literature gap by conducting a Monte Carlo study.

Model-based and design-based multilevel approaches

Two types of analytical approaches are widely employed for analyzing clustered longitudinal data, namely the model-based

approach and the design-based approach. The *model-based approach* specifies level-specific models on the basis of the data structure (Heck & Thomas, 2009; B. O. Muthén & Asparouhov, 2011). Within the framework of multilevel structural equation modeling (MSEM), the time dimension can be converted into a multivariate vector and, therefore, the three-level clustered longitudinal data can be analyzed with a two-level model (i.e., Multilevel Latent Growth Curve Model – MLGCM; Bovaird, 2007; Heck & Thomas, 2009; B. O. Muthén & Asparouhov, 2009), where the intraindividual growth and interindividual variability is modeled at the *within-level model* and the cluster-level growth is modeled at the *between-level model* (see for example, Duncan et al., 1997). If researchers are interested in only the individual-level growth trajectory (e.g., psychological virtues during adolescence; Ferragut, Blanca, & Ortiz-Tallo, 2014), then the parameter estimates in the within-level model of MLGCM (see Fig. 1) can theoretically fulfill researchers' needs.

In addition to MLGCM, another model-based technique within the framework of MSEM, namely the two-level maximum model (MM; Hox, 2010), might be an alternative to studying clustered longitudinal data. MM (see Fig. 2) is often preferred given that researchers are primarily interested in individual-level growth trajectory. In particular, MM specifies the target/hypothesized growth pattern in the within-level model only and saturates the between-level model. The performance of MM has been examined using two-level measurement models in previous studies, which indicated that MM was capable of providing unbiased fixed effect estimates, variances components, and accurate *SEs* in the within-level model (Pornprasertmanit et al., 2014; Wu & Kwok, 2012). Moreover, Wu and Kwok (2012) suggested that MM can be used as a substitute for MLGCM, especially when sample sizes are small (e.g., < 50). However, to the best of our knowledge, no studies have been conducted to investigate whether MM demonstrates promise in analyzing individual growth with clustered longitudinal data. In the present study, we hypothesized that MM could generate accurate parameter estimates and *SEs* of intraindividual growth and interindividual variability, similar to those derived by MLGCM.

The *design-based approach* takes into account clustering in longitudinal data by adjusting the *SEs* of parameter estimates (i.e., robust *SEs*) using a sandwich estimator based on the sampling design and, therefore, it produces accurate statistical inferences (Heeringa, West, & Berglund, 2010; L. K. Muthén & Muthén, 1998–2015; Stapleton, 2006, 2008). In *Mplus*, a design-based latent growth curve model (D-LGCM) can be implemented by specifying a model exactly like the within-level model of MLGCM presented in Fig. 1 and using the command “TYPE = COMPLEX” to adjust *SEs* of parameter estimates in the model (L. K. Muthén & Muthén, 1998–2015). Similar to MM, D-LGCM might be favored if the primary interest is the individual-level growth trajectory.

The performance of D-LGCM has been examined by Wu et al. (2015) in terms of the accuracy of the regression coefficients of

Fig. 1 A two-level multilevel latent growth curve model

between-level and within-level predictors. More specifically, Wu and colleagues generated a simulated dataset based on a conditional MLGCM – one predictor at the between-level and within-level, respectively. Their simulation results suggested the D-LGCM, including predictors from both levels, could produce unbiased regression coefficients and *SEs*. However, whether a D-LGCM can produce accurate intraindividual growth and interindividual variability has not been well investigated.

In summary, researchers need be aware that models ignoring individual dependency should not be used for analyzing clustered longitudinal data. Researchers who are primarily interested in the individual-level growth trajectory have two model-based approaches (MLGCM and MM) and one design-based approach (D-LGCM) as options to analyze clustered longitudinal data. Until now, understanding the viability of MM as an alternative to analyze clustered longitudinal data has not been well studied.

Fig. 2 A two-level maximum model

Moreover, although D-LGCM is theoretically expected to provide robust *SEs* for statistical inferences in the individual-level model, the extent to which D-LGCM can produce unbiased parameter estimates and *SEs* of intraindividual growth and interindividual variability is still unknown.

The purpose of the present study was to contribute to the literature gap by examining the effectiveness of two model-based approaches (MLGCM and MM) and one design-based approach (D-LGCM), which have the potential to accommodate individual dependency in clustered longitudinal data. Simulated data (1,000 replications) were generated based on the findings of empirical studies of the mathematics literacy trajectory derived from clustered longitudinal data (Baumert, Nagy, & Lehmann, 2012). We analyzed the simulated data with four approaches – MLGCM, MM, D-LGCM as well as a model failing to accommodate individual dependency (Single-level Latent Growth Curve Model – SLGCM). The model solutions of SLGCM presented the consequences of ignoring individual dependency when clustered longitudinal data was inappropriately analyzed. When generating the simulated data, we manipulated two design factors, namely, number of clusters and cluster size, to evaluate the impact of sample size at different levels on the performance of the four approaches. Parameter estimates and *SEs* produced by MLGCM were expected to be less biased. The effectiveness of MM, D-LGCM, as well as SLGCM was examined in terms of unbiasedness and efficiency of parameter estimates in the marginal mean structure (i.e., intraindividual growth) and covariance structure (i.e., interindividual variability).

Method

A Monte Carlo study was conducted using *Mplus* 7.31 (L. K. Muthén & Muthén, 1998–2015) to evaluate the performance of two model-based approaches (MLGCM and MM), a design-based approach (D-LGCM), and a model failing to accommodate individual dependency (SLGCM). A MLGCM presented in Fig. 1 was used to generate simulated clustered longitudinal data. In order to increase the generalizability of our findings, parameter settings were adopted from findings in an empirical study (Baumert et al., 2012), which investigated fourth graders' growth trajectory of mathematics literacy scores. Two design factors were considered in our simulation: number of clusters (NC) and cluster size (CS). Simulated data were then analyzed by MLGCM, MM, D-LGCM, and SLGCM. *Mplus* syntax for the present simulation study are provided in Table 5 of Appendix.

Population model

According Kwok, West, and Green's (2007) review of longitudinal studies published in the journal *Developmental Psychology* in 2002, the average time point of measurement was 4. Therefore,

a population MLGCM model with four measurement time points was used to create our simulated data. Figure 1 presents the population MLGCM model with parameter settings adopted from Baumert et al.'s (2012) study. The repeated measures were denoted as Y1 – Y4. The intercepts of the repeated measures were fixed at zero. The linear growth pattern was modeled in the within- and between-level models. The factor loadings of the intercept factors (IW and IB) were fixed at 1.0 and those of the linear slope factors (SW and SB) were set as 0, 1, 2, and 3 as part of the growth model parameterization. The parameter settings of the marginal mean structure and covariance structure in the within-level model are presented in the following matrices. In the marginal mean structure matrix (α_w), the means of IW and SW are fixed to zero, and referred to as dummy zero means by Muthén (1997). Note the means of IW and SW are estimated in the between-level model (B. O. Muthén, 1997).

In the covariance structure matrix (Φ_w), the diagonals are the variances of IW (113.30; π_{00}) and SW (8.77; π_{11}), where π_{00} denotes the interindividual variability of initial mathematics literacy scores and π_{11} signifies the interindividual variability of the linear trajectories. The off-diagonals in the Φ_w are the covariance between IW and SW ($\pi_{10} = 0.315$; can be converted into a correlation equal to .01), which suggests that the relationship between individual initial scores and growth trajectories was trivial. The error variances of the outcome variables were constrained to be equal over time and set to 40.57. In other words, the population model showed students' initial mathematics literacy scores (π_{00}) and linear growth in mathematics literacy (π_{11}) varied across individuals but students' initial performance was only trivially related to later growth (small π_{10}).

$$\alpha_w = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Phi_w = \begin{bmatrix} 1130(\pi_{00}) & 0.315 \\ 0.315(\pi_{10}) & 8.77(\pi_{11}) \end{bmatrix}$$

The structure of the between-level model was identical to that of the within-level model in our simulation. The parameter settings of the marginal mean structure and covariance structure in between-level models are presented in the matrices α_B and Φ_B , respectively. The means of IB (α_1) and SB (α_2) are set to 96.28 and 9.94, respectively, where α_1 denotes average initial mathematics literacy score and α_2 signifies the average linear trajectory of mathematics literacy scores.

In the covariance structure matrix (Φ_B), the diagonals are the variances of IB (28.86; γ_{00}) and SB (2.80; γ_{11}), where γ_{00} denotes the variability of initial mathematics literacy scores across cluster units (e.g., school) and γ_{11} signifies the variability of the linear trajectory across cluster units. The off-diagonals in the Φ_B are the covariance between IB and SB ($\gamma_{10} = 2.607$; correlation = .29), which suggests that the cluster-level initial mathematics literacy scores were positively related to cluster-level linear trajectories. The errors variances of the outcome variables were set to 0.

$$\alpha_B = \begin{bmatrix} 96.28(\alpha_1) \\ 9.94(\alpha_2) \end{bmatrix}, \Phi_B = \begin{bmatrix} 28.86(\gamma_{00}) & 2.607 \\ 2.607(\gamma_{10}) & 2.80(\gamma_{11}) \end{bmatrix}$$

Note the current study intended to evaluate to what extent three multilevel approaches were able to produce unbiased estimates of intraindividual growth (α_1 and α_2) and interindividual variability (π_{00} , π_{11} , and π_{10}) and the corresponding *SEs*. Note π_{00} denotes the interindividual variability of individual initial scores, π_{11} signifies the interindividual variability of the linear trajectories, and π_{10} is the relationship between individual initial scores and growth trajectories. The parameter settings of γ_{00} , γ_{11} , and γ_{10} were required for simulated data generation but were not the focus of the study. Given the parameter settings listed above, the intraclass correlations for Y1, Y2, Y3, and Y4 ranged from 0.231 to 0.299, which indicated non-ignorable individual dependency (Hox, 2010; B. O. Muthén & Satorra, 1995) in the simulated data.

Design factors

Design factors considered in the present simulation study were as follows: number of clusters (NC = 10, 30, 50, 100, 150, 200, and 500) and cluster size (CS = 5, 10, and 20). An extremely small NC (10) was specifically considered to evaluate the boundaries of performance of MLGCM, MM, and D-LGCM. On the other hand, we also considered a large NC condition (500) because some current large-scale longitudinal studies have collected data from national representative samples who were nested within a tremendous number of clusters. For example, in the base year of the Educational Longitudinal Study of 2002 (ELS: 2002), 15,362 students nested within 752 schools were interviewed in 2002 and followed-up in the following 10 years (Ingels, Pratt, Rogers, Siegel, & Stutts, 2005). Therefore, including a large NC condition in our simulation design can increase the generalizability of our findings to large-scale clustered longitudinal data analyses. Moreover, because CS was less likely to create biased estimates and *SEs* in MSEM (Hox & Maas, 2001), we adopted small, medium, and large values of CS, which were commonly found in substantive research.

There were a total of $7 \text{ (NC)} \times 3 \text{ (CS)} = 21$ conditions. For each condition, replications with convergence problems or improper solutions (e.g., negative unique variances) were excluded until at least 1,000 replications were generated. To compute the empirical power of each key parameter value in the population model a Monte Carlo approach (L. K. Muthén & Muthén, 2002) was applied (see Table 6 of Appendix). The covariance between IW and SW at the within-level (π_{10}) exhibited minimal power ($\leq .106$) across all simulation conditions; while covariance between IB and SB at the between-level (γ_{10}) exhibited power larger than .80 when (a) NC = 200 and CS = 20 or (b) NC = 500, regardless of CS. Other parameters had reasonable power across all simulation conditions, except for NC = 30 and CS = 5 condition. Each set of

replicates was analyzed with four approaches – MLGCM, MM, D-LGCM, and SLGCM.

Relative biases

Because parameter estimates and estimated *SEs* in the current study were not in standardized form, relative biases (Hoogland & Boomsma, 1998) were computed to evaluate the accuracy of model solutions derived through MLGCM, MM, D-LGCM, and SLGCM. The formulas for computing relative biases for parameter estimates (θ) and estimated *SEs* are presented as follows:

$$\text{Relative bias } (\theta) = \frac{\bar{\theta}_{est} - \theta_{true}}{\theta_{true}} \quad (1)$$

$$\text{Relative bias } (SE_{\theta}) = \frac{\overline{SE}_{\theta_{est}} - \sigma_{\theta_{est}}}{\sigma_{\theta_{est}}} \quad (2)$$

where $\bar{\theta}_{est}$ is the average of parameter estimates in each condition, θ_{true} is the population value of the parameter, $\overline{SE}_{\theta_{est}}$ is the average *SE* of a parameter estimate, and $\sigma_{\theta_{est}}$ is the standard deviation of the parameter estimate across successful replications. Note that relative bias is an average measure of accuracy for each parameter estimate (θ) and estimated *SE*. According to Hoogland and Boomsma (1998), relative bias with an absolute value larger than 0.05 is considered unacceptable. In the current study, an estimate with an absolute value of relative bias less than 0.01 was deemed fairly accurate. If needed, factorial ANOVAs were conducted to determine the contributions of the design factors to estimates of parameters and *SEs*. The total sum of squares provided the variability of the parameter and *SE* estimates across all successful replications while eta-squared (η^2) indicated the proportion of the variance accounted for by a particular design factor or the interaction effect term. Notably, η^2 was obtained by dividing the Type III sum of squares of a particular predictor or the interaction effect by the corrected total sum of squares.

Results

Convergence rate/improper solution of simulations

Results suggested that when sample size at the between-level was too small (NC = 10), applying MLGCM or MM could lead to serious convergence problems (convergence rate close to 0) and improper solutions (e.g., negative unique variances), regardless of varying CS conditions. On the other hand, satisfactory convergence rates and proper solutions were found in other sample size settings (NC = 30, 50, 100, 150, 200, 500; CS = 5, 10, 20) when

Table 1 Parameter settings in mean structure of population model and parameter estimates using different approaches

		α_1 (mean of intercept factor = 96.28)				α_2 (mean of slope factor = 9.94)			
		MLGCM							
NC	CS	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	96.20	0.00	1.34	-0.03	9.94	0.00	0.44	-0.02
30	10	96.26	0.00	1.17	-0.03	9.94	0.00	0.38	-0.03
30	20	96.27	0.00	1.07	-0.02	9.96	0.00	0.34	-0.03
50	5	96.23	0.00	1.05	0.02	9.95	0.00	0.35	0.00
50	10	96.26	0.00	0.92	0.01	9.94	0.00	0.30	0.00
50	20	96.30	0.00	0.84	-0.01	9.96	0.00	0.27	-0.04
100	5	96.28	0.00	0.75	0.00	9.94	0.00	0.25	0.00
100	10	96.27	0.00	0.65	0.03	9.93	0.00	0.21	0.00
100	20	96.29	0.00	0.60	0.02	9.95	0.00	0.19	0.00
150	5	96.28	0.00	0.61	0.00	9.94	0.00	0.20	0.00
150	10	96.27	0.00	0.53	0.06	9.93	0.00	0.17	0.00
150	20	96.30	0.00	0.49	0.00	9.95	0.00	0.15	-0.06
200	5	96.27	0.00	0.53	-0.02	9.94	0.00	0.18	0.00
200	10	96.27	0.00	0.46	0.02	9.93	0.00	0.15	0.00
200	20	96.29	0.00	0.42	0.05	9.95	0.00	0.13	0.00
500	5	96.27	0.00	0.34	0.03	9.94	0.00	0.11	0.00
500	10	96.28	0.00	0.29	0.04	9.94	0.00	0.09	0.00
500	20	96.30	0.00	0.27	0.00	9.94	0.00	0.08	0.00
MM									
NC	CS	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	96.21	0.00	1.46	0.06	9.95	0.00	0.48	0.07
30	10	96.27	0.00	1.24	0.03	9.93	0.00	0.40	0.03
30	20	96.31	0.00	1.13	0.04	9.97	0.00	0.38	0.09
50	5	96.24	0.00	1.09	0.06	9.95	0.00	0.35	0.00
50	10	96.26	0.00	0.92	0.01	9.93	0.00	0.30	0.00
50	20	96.29	0.00	0.84	-0.01	9.96	0.00	0.27	-0.04
100	5	96.28	0.00	0.75	0.00	9.94	0.00	0.25	0.00
100	10	96.27	0.00	0.65	0.03	9.93	0.00	0.21	0.00
100	20	96.29	0.00	0.59	0.00	9.95	0.00	0.19	0.00
150	5	96.28	0.00	0.61	0.00	9.94	0.00	0.20	0.00
150	10	96.27	0.00	0.53	0.06	9.93	0.00	0.17	0.00
150	20	96.30	0.00	0.49	0.00	9.95	0.00	0.15	-0.06
200	5	96.29	0.00	0.53	-0.02	9.94	0.00	0.18	0.00
200	10	96.27	0.00	0.46	0.02	9.93	0.00	0.15	0.00
200	20	96.29	0.00	0.42	0.05	9.95	0.00	0.13	0.00
500	5	96.27	0.00	0.34	0.03	9.94	0.00	0.11	0.00
500	10	96.29	0.00	0.29	0.04	9.94	0.00	0.09	0.00
500	20	96.30	0.00	0.27	0.00	9.94	0.00	0.08	0.00
D-LGCM									
NC	CS	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	96.21	0.00	1.37	-0.01	9.94	0.00	0.45	0.00
30	10	96.26	0.00	1.19	-0.01	9.94	0.00	0.38	-0.03
30	20	96.29	0.00	1.09	0.00	9.95	0.00	0.35	0.00
50	5	96.23	0.00	1.06	0.03	9.95	0.00	0.35	0.00
50	10	96.26	0.00	0.93	0.02	9.94	0.00	0.30	0.00
50	20	96.30	0.00	0.84	-0.01	9.96	0.00	0.27	-0.04

Table 1 (continued)

		α_1 (mean of intercept factor = 96.28)				α_2 (mean of slope factor = 9.94)			
		MLGCM							
100	5	96.28	0.00	0.75	0.00	9.94	0.00	0.25	0.00
100	10	96.27	0.00	0.66	0.05	9.93	0.00	0.21	0.00
100	20	96.29	0.00	0.60	0.02	9.95	0.00	0.19	0.00
150	5	96.27	0.00	0.62	0.02	9.94	0.00	0.20	0.00
150	10	96.27	0.00	0.53	0.06	9.93	0.00	0.17	0.00
150	20	96.30	0.00	0.49	0.00	9.95	0.00	0.16	0.00
200	5	96.27	0.00	0.53	-0.02	9.94	0.00	0.18	0.00
200	10	96.27	0.00	0.46	0.02	9.93	0.00	0.15	0.00
200	20	96.29	0.00	0.42	0.05	9.95	0.00	0.13	0.00
500	5	96.27	0.00	0.34	0.03	9.94	0.00	0.11	0.00
500	10	96.28	0.00	0.29	0.04	9.94	0.00	0.09	0.00
500	20	96.30	0.00	0.27	0.00	9.94	0.00	0.08	0.00
		SLGCM							
NC	CS	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	96.21	0.00	1.06	0.03	9.94	0.00	0.36	0.03
30	10	96.28	0.00	0.75	-0.18	9.94	0.00	0.25	-0.17
30	20	96.30	0.00	0.53	-0.38	9.97	0.00	0.18	-0.36
50	5	96.23	0.00	0.82	-0.20	9.95	0.00	0.28	-0.20
50	10	96.26	0.00	0.59	-0.35	9.94	0.00	0.20	-0.33
50	20	96.30	0.00	0.41	-0.52	9.96	0.00	0.14	-0.50
100	5	96.28	0.00	0.58	-0.23	9.94	0.00	0.20	-0.20
100	10	96.27	0.00	0.41	-0.35	9.93	0.00	0.14	-0.33
100	20	96.29	0.00	0.29	-0.51	9.95	0.00	0.10	-0.47
150	5	96.27	0.00	0.48	-0.21	9.94	0.00	0.16	-0.20
150	10	96.27	0.00	0.34	-0.32	9.93	0.00	0.11	-0.35
150	20	96.30	0.00	0.24	-0.51	9.95	0.00	0.08	-0.50
200	5	96.27	0.00	0.41	-0.24	9.94	0.00	0.14	-0.22
200	10	96.27	0.00	0.29	-0.36	9.93	0.00	0.10	-0.33
200	20	96.29	0.00	0.21	-0.48	9.95	0.00	0.07	-0.46
500	5	96.27	0.00	0.26	-0.21	9.94	0.00	0.09	-0.18
500	10	96.28	0.00	0.18	-0.36	9.94	0.00	0.06	-0.33
500	20	96.30	0.00	0.13	-0.52	9.94	0.00	0.04	-0.50

Note. NC number of cluster, CS cluster size, *Est.* parameter estimate, *SE* standard error of parameter estimate, *Bias* relative bias ($|\text{relative bias}| \geq 0.05$ is considered unacceptable and highlighted in bold type), *MLGCM* multilevel latent growth curve model, *MM* maximum model, *D-LGCM* design-based latent growth curve model, *SLGCM* single-level latent growth curve model

MLGCM and MM were applied to analyze simulated data (above 98.61%). D-LGCM and SLGCM performed similarly to one another and both had convergence rates equal to 100% and proper solutions across all sample size settings. Next we present the results of parameter estimates and *SEs* under the various simulation conditions, excluding NC = 10.

Parameter estimates and *SEs* in the mean structure

Table 1 presents the parameter estimates and *SEs* for two parameters (α_1 and α_2) as well as their corresponding relative

biases in the mean structure derived from four approaches, where α_1 denotes average initial mathematics literacy score and α_2 signifies the average linear trajectory of mathematics literacy scores. Note that MLGCM estimated parameters in the mean structure at the between-level, while the other three approaches estimated parameters at the within-level. We found MLGCM was capable of generating fairly accurate estimates of the intercept factor mean (α_1 ; ranging from 96.20 to 96.30; $|\text{relative biases}| = 0.00$) and slope factor mean (α_2 ; ranging from 9.93 to 9.96; $|\text{relative biases}| = 0.00$) across all sample size scenarios. Moreover, all *SEs* derived from

Table 2 Parameter settings in the covariance structure of the population within-level model and parameter estimates and *SEs* using different approaches

NC	CS	π_{00} (variance of intercept factor = 113.30)				π_{11} (variance of slope factor = 8.77)				π_{10} (covariance between two factors = 0.32)			
		MLGCM											
		Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	112.41	-0.01	17.94	-0.06	8.89	0.01	2.49	-0.05	0.77	1.41	4.90	-0.04
30	10	112.46	-0.01	12.21	-0.03	8.73	0.00	1.71	-0.03	0.42	0.31	3.35	-0.06
30	20	112.81	0.00	8.40	-0.03	8.77	0.00	1.19	-0.06	0.42	0.31	2.32	-0.08
50	5	113.17	0.00	14.20	-0.02	8.71	-0.01	1.96	-0.02	0.42	0.31	3.87	-0.02
50	10	112.78	0.00	9.60	-0.02	8.77	0.00	1.34	-0.02	0.30	-0.06	2.63	-0.01
50	20	113.32	0.00	6.63	-0.03	8.77	0.00	0.93	-0.05	0.36	0.13	1.82	-0.08
100	5	113.17	0.00	10.22	0.00	8.72	-0.01	1.41	-0.02	0.36	0.13	2.77	-0.02
100	10	113.05	0.00	6.88	-0.01	8.75	0.00	0.96	-0.03	0.35	0.09	1.87	-0.02
100	20	113.17	0.00	4.72	0.00	8.76	0.00	0.66	-0.03	0.39	0.22	1.29	-0.07
150	5	113.24	0.00	8.39	-0.01	8.73	0.00	1.16	-0.02	0.40	0.25	2.27	-0.03
150	10	113.14	0.00	5.62	0.02	8.75	0.00	0.78	-0.03	0.33	0.03	1.53	-0.04
150	20	113.11	0.00	3.87	0.00	8.76	0.00	0.54	-0.04	0.35	0.09	1.06	-0.09
200	5	113.37	0.00	7.29	0.00	8.75	0.00	1.00	0.01	0.40	0.25	1.98	-0.02
200	10	113.16	0.00	4.87	0.02	8.77	0.00	0.68	0.00	0.36	0.13	1.33	-0.04
200	20	113.16	0.00	3.36	0.01	8.76	0.00	0.47	-0.04	0.36	0.13	0.92	-0.08
500	5	113.12	0.00	4.61	-0.02	8.76	0.00	0.64	0.02	0.35	0.09	1.25	-0.05
500	10	113.17	0.00	3.09	-0.02	8.76	0.00	0.43	0.00	0.34	0.06	0.84	-0.07
500	20	113.25	0.00	2.13	0.03	8.77	0.00	0.30	-0.03	0.33	0.03	0.58	-0.03
NC	CS	MM											
		Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	112.51	-0.01	17.72	-0.07	8.85	0.01	2.50	-0.05	0.35	0.09	4.86	-0.04
30	10	113.21	0.00	12.14	-0.04	8.83	0.01	1.70	-0.04	0.19	-0.41	3.35	-0.06
30	20	113.22	0.00	8.44	-0.02	8.84	0.01	1.19	-0.06	0.29	-0.09	2.31	-0.08
50	5	113.57	0.00	14.13	-0.03	8.88	0.01	1.97	-0.02	0.19	-0.41	3.87	-0.02
50	10	113.13	0.00	9.59	-0.02	8.88	0.01	1.35	-0.01	0.13	-0.59	2.63	-0.01
50	20	113.52	0.00	6.63	-0.03	8.83	0.01	0.93	-0.05	0.27	-0.16	1.82	-0.08
100	5	113.62	0.00	10.22	0.00	8.88	0.01	1.43	-0.01	0.14	-0.56	2.79	-0.02
100	10	113.34	0.00	6.88	-0.01	8.83	0.01	0.97	-0.02	0.22	-0.31	1.88	-0.02
100	20	113.31	0.00	4.72	0.00	8.80	0.00	0.67	-0.01	0.33	0.03	1.29	-0.07
150	5	113.67	0.00	8.40	-0.01	8.86	0.01	1.17	-0.01	0.20	-0.38	2.29	-0.02
150	10	113.40	0.00	5.62	0.02	8.82	0.01	0.79	-0.01	0.22	-0.31	1.54	-0.04
150	20	113.23	0.00	3.87	0.00	8.80	0.00	0.55	-0.02	0.30	-0.06	1.06	-0.09
200	5	113.77	0.00	7.33	0.01	8.87	0.01	1.02	0.03	0.22	-0.31	1.99	-0.01
200	10	113.37	0.00	4.88	0.02	8.83	0.01	0.69	0.01	0.27	-0.16	1.34	-0.03
200	20	113.26	0.00	3.36	0.01	8.79	0.00	0.47	-0.04	0.31	-0.03	0.92	-0.08
500	5	113.43	0.00	4.62	-0.02	8.85	0.01	0.65	0.03	0.22	-0.31	1.27	-0.03
500	10	113.32	0.00	3.09	-0.02	8.80	0.00	0.44	0.02	0.28	-0.13	0.89	-0.01
500	20	113.32	0.00	2.13	0.03	8.79	0.00	0.30	-0.03	0.30	-0.06	0.58	-0.03
NC	CS	D-LGCM											
		Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	140.42	0.24	20.57	0.08	11.44	0.30	2.68	0.02	3.15	8.84	5.42	0.07
30	10	140.49	0.24	15.45	0.23	11.37	0.30	1.97	0.11	2.86	7.94	3.99	0.12
30	20	141.00	0.24	12.00	0.39	11.47	0.31	1.49	0.17	2.81	7.78	3.01	0.19
50	5	140.98	0.24	16.10	0.11	11.48	0.31	2.10	0.04	3.02	8.44	4.22	0.07
50	10	141.13	0.25	12.13	0.24	11.49	0.31	1.54	0.12	2.90	8.06	3.13	0.18
50	20	141.58	0.25	9.45	0.39	11.50	0.31	1.16	0.18	2.87	7.97	2.37	0.20
100	5	141.13	0.25	11.52	0.13	11.50	0.31	1.50	0.04	3.01	8.41	3.00	0.06

Table 2 (continued)

100	10	141.74	0.25	8.70	0.25	11.54	0.32	1.10	0.11	2.97	8.28	2.23	0.17
100	20	141.76	0.25	6.74	0.42	11.51	0.31	0.82	0.21	2.93	8.16	1.69	0.22
150	5	141.56	0.25	9.46	0.12	11.52	0.31	1.23	0.04	3.09	8.66	2.46	0.06
150	10	141.77	0.25	7.11	0.29	11.52	0.31	0.90	0.13	2.94	8.19	1.82	0.14
150	20	141.72	0.25	5.52	0.42	11.52	0.31	0.67	0.20	2.90	8.06	1.38	0.19
200	5	141.69	0.25	8.21	0.13	11.53	0.31	1.07	0.08	3.05	8.53	2.13	0.05
200	10	141.90	0.25	6.17	0.29	11.54	0.32	0.78	0.15	2.97	8.28	1.58	0.14
200	20	141.77	0.25	4.79	0.43	11.53	0.31	0.58	0.18	2.91	8.09	1.20	0.20
500	5	141.83	0.25	5.21	0.10	11.54	0.32	0.68	0.08	2.97	8.28	1.35	0.03
500	10	141.95	0.25	3.90	0.24	11.56	0.32	0.50	0.16	2.93	8.16	1.00	0.11
500	20	142.01	0.25	3.05	0.48	11.55	0.32	0.37	0.19	2.92	8.13	0.76	0.27
NC	CS	SLGCM											
		Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	140.33	0.24	20.08	0.05	11.32	0.29	2.68	0.02	3.12	8.75	5.26	0.04
30	10	141.08	0.25	14.25	0.13	11.41	0.30	1.90	0.07	2.93	8.16	3.73	0.05
30	20	140.84	0.24	10.06	0.16	11.41	0.30	1.34	0.06	2.84	7.88	2.64	0.05
50	5	140.98	0.24	15.59	0.07	11.48	0.31	2.08	0.03	3.02	8.44	4.08	0.04
50	10	141.13	0.25	11.04	0.13	11.49	0.31	1.47	0.07	2.90	8.06	2.89	0.09
50	20	141.58	0.25	7.83	0.15	11.50	0.31	1.04	0.06	2.87	7.97	2.05	0.04
100	5	141.13	0.25	11.04	0.08	11.50	0.31	1.48	0.03	3.01	8.41	2.90	0.02
100	10	141.74	0.25	7.83	0.13	11.54	0.32	1.04	0.05	2.97	8.28	2.05	0.07
100	20	141.76	0.25	5.54	0.17	11.51	0.31	0.74	0.09	2.93	8.16	1.45	0.05
150	5	141.56	0.25	9.04	0.07	11.52	0.31	1.21	0.03	3.09	8.66	2.37	0.02
150	10	141.77	0.25	6.40	0.16	11.52	0.31	0.85	0.06	2.94	8.19	1.67	0.04
150	20	141.72	0.25	4.52	0.16	11.52	0.31	0.60	0.07	2.90	8.06	1.18	0.02
200	5	141.69	0.25	7.83	0.08	11.53	0.31	1.04	0.05	3.05	8.53	2.05	0.01
200	10	141.90	0.25	5.54	0.16	11.54	0.32	0.74	0.09	2.97	8.28	1.45	0.05
200	20	141.77	0.25	3.92	0.17	11.53	0.31	0.52	0.06	2.91	8.09	1.03	0.03
500	5	141.83	0.25	4.96	0.05	11.54	0.32	0.66	0.05	2.97	8.28	1.30	-0.01
500	10	141.95	0.25	3.51	0.12	11.56	0.32	0.47	0.09	2.93	8.16	0.92	0.02
500	20	142.01	0.25	2.48	0.20	11.55	0.32	0.33	0.06	2.92	8.13	0.65	0.08

Note. NC number of cluster, CS cluster size, *Est.* parameter estimate, *SE* standard error of parameter estimate, *Bias* relative bias ($|\text{relative bias}| \geq 0.05$ is considered unacceptable and highlighted in bold type), *MLGCM* multilevel latent growth curve model, *MM* maximum model, *D-LGCM* design-based latent growth curve model, *SLGCM* single-level latent growth curve model

MLGCM were not biased except for a few trivially biased ones ($|\text{relative biases}|$ ranged from 0.00 to 0.06) across all sample size scenarios.

Likewise, we also found MM and D-LGCM generated unbiased parameter estimates of α_1 and α_2 ($|\text{relative biases}| = 0.00$) and unbiased corresponding *SEs* except for a few trivially biased ones ($|\text{relative biases}|$ ranged from 0.00 to 0.09) across all sample size scenarios. That is, these two approaches performed in the same way as MLGCM did. When SLGCM was applied, the parameter estimates of α_1 and α_2 were unbiased ($|\text{relative biases}| = 0.00$), but their *SEs* were substantially downwardly biased ($|\text{relative biases}|$ ranged from 0.03 to 0.52).

Parameter estimates and *SEs* in the covariance structure of the within-level model

Table 2 presents the parameter estimates and *SEs* for three parameters in the within-level covariance structure (π_{00} , π_{11} , and π_{10}) derived from four approaches, where π_{00} denotes interindividual variability of initial mathematics literacy scores, π_{11} signifies interindividual variability of the linear trajectories, and π_{10} represents the relationship between initial

scores and trajectories. Additionally, the corresponding relative biases for each parameter estimate and *SE* are also presented. When MLGCM was applied, intercept factor variance (π_{00} ; ranging from 112.41 to 113.37) as well as slope factor variance (π_{11} ; ranging from 8.71 to 8.89) were accurately estimated (i.e., $|\text{relative bias}|$ ranged from 0 to 0.01) across all sample size scenarios. The estimates for the covariance between intercept and slope (π_{10} ; ranging from 0.30 to 0.77) tended to be positively biased ($|\text{relative bias}|$ ranged from 0.03 to 1.41). Our further exploration with ANOVA suggested the estimates of π_{10} were trivially explained by the main effects of NC and CS and the interaction effect (η^2 close to 0). In addition, we found all $|\text{relative biases}|$ of *SEs* were less than 0.05 except for a few trivially biased *SEs* (ranging from 0.05 to 0.09). Generally speaking, the results suggested reasonable accuracy of *SEs*.

MM performed similarly to MLGCM except for the parameter estimates of π_{10} . Results showed reasonably accurate estimates of π_{00} (ranging from 112.51 to 113.77), π_{11} (ranging from 8.79 to 8.88) and *SEs* ($|\text{relative biases}| \leq 0.06$). However, π_{10} tended to be underestimated (ranging from 0.13 to 0.35) across all sample size scenarios with $|\text{relative biases}|$ ranging

Table 3 Parameter settings in the covariance structure of the population between-level model and parameter estimates using MLGCM

		γ_{00} (variance of intercept factor = 28.86)				γ_{11} (variance of slope factor = 2.80)				γ_{10} (covariance between two factors = 2.61)			
MLGCM													
NC	CS	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias	Est.	Est. Bias	SE	SE Bias
30	5	27.64	-0.04	14.17	0.00	2.76	-0.01	1.56	-0.05	2.49	-0.05	3.43	0.06
30	10	28.13	-0.03	10.28	-0.08	2.69	-0.04	1.07	-0.04	2.52	-0.03	2.42	-0.07
30	20	27.95	-0.03	8.51	-0.12	2.71	-0.03	0.86	-0.08	2.51	-0.04	1.97	-0.09
50	5	27.85	-0.03	11.08	-0.01	2.77	-0.01	1.21	-0.03	2.59	-0.01	2.65	0.04
50	10	28.40	-0.02	8.22	-0.03	2.73	-0.02	0.86	-0.04	2.59	-0.01	1.91	-0.05
50	20	28.28	-0.02	6.81	-0.08	2.73	-0.02	0.69	0.00	2.50	-0.04	1.57	-0.05
100	5	27.99	-0.03	8.03	-0.01	2.78	-0.01	0.88	-0.02	2.65	0.02	1.90	0.03
100	10	28.73	0.00	6.00	-0.03	2.80	0.00	0.63	0.02	2.61	0.00	1.38	-0.01
100	20	28.60	-0.01	4.95	-0.06	2.75	-0.02	0.50	0.00	2.54	-0.03	1.13	-0.05
150	5	28.33	-0.02	6.64	0.00	2.79	0.00	0.73	0.04	2.69	0.03	1.57	0.02
150	10	28.65	-0.01	4.93	0.02	2.77	-0.01	0.51	0.02	2.60	0.00	1.13	-0.01
150	20	28.61	-0.01	4.06	-0.02	2.76	-0.01	0.41	0.02	2.54	-0.03	0.93	-0.04
200	5	28.32	-0.02	5.78	0.01	2.78	-0.01	0.63	0.03	2.66	0.02	1.36	-0.01
200	10	28.75	0.00	4.29	0.00	2.77	-0.01	0.44	0.00	2.60	0.00	0.98	-0.02
200	20	28.61	-0.01	3.53	-0.03	2.77	-0.01	0.36	0.00	2.55	-0.02	0.81	-0.04
500	5	28.75	0.00	3.70	0.01	2.79	0.00	0.40	0.00	2.61	0.00	0.87	-0.02
500	10	28.79	0.00	2.73	0.02	2.80	0.00	0.28	0.04	2.59	-0.01	0.63	0.02
500	20	28.75	0.00	2.26	-0.03	2.78	-0.01	0.23	0.00	2.59	-0.01	0.52	0.00

Note. NC number of cluster, CS cluster size, Est. parameter estimate, SE standard error of parameter estimate, Bias relative bias ($|\text{relative bias}| \geq 0.05$ is considered unacceptable and highlighted in bold type), MLGCM multilevel latent growth curve model

from 0.03 to 0.59. Further ANOVA explorations suggested the variation of estimated π_{10} was not explained by the main effects of NC and CS and interaction effect (η^2 close to 0).

Both D-LGCM and SLGCM approaches produced similar and severely biased parameter estimates of π_{00} , π_{11} , and π_{10} . For example, the parameter estimates of within-level intercept factor variance (π_{00}) derived from the two approaches ranged from 140.42 to 142.01 and were upwardly biased ($|\text{relative biases}|$ ranged from 0.24 to 0.25). It should be noted that the estimated within-level intercept factor variance was

approximately equal to the sum of the population intercept factor variance within level (113.30) and between level (28.86). Apparently, intercept factor variance in the between-level model was redistributed to the within-level model. Similarly, the within-level slope factor variance (π_{11} ranged from 11.32 to 11.56) estimated by both approaches was also upwardly biased ($|\text{relative biases}|$ ranged from 0.29 to 0.32) and close to the sum of the population slope factor variance in the within-level (8.77) and between-level (2.80) models. In the same manner, the within-level covariance between

Table 4 Results of an empirical example

	Mean structure				Covariance structure of within-level model						Covariance structure of between-level model					
	Mean of intercept factor		Mean of slope factor		Variance of intercept factor		Variance of slope factor		Covariance between two factors		Variance of intercept factor		Variance of slope factor		Covariance between two factors	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
MLGCM	49.95	0.62	4.07	0.12	70.57	3.32	3.83	0.30	6.42	0.54	17.31	3.33	0.32	0.12	1.60	0.51
MM	49.78	0.60	4.04	0.12	70.60	3.32	3.83	0.30	6.41	0.54	-	-	-	-	-	-
D-LGCM	50.12	0.64	4.02	0.11	87.37	4.62	4.08	0.34	8.03	0.59	-	-	-	-	-	-
SLGCM	50.12	0.18	4.02	0.05	87.37	2.58	4.08	0.22	8.03	0.55	-	-	-	-	-	-

Note. MLGCM multilevel latent growth curve model, MM maximum model, D-LGCM design-based latent growth curve model, SLGCM single-level latent growth curve model

intercept and slope (π_{10}) estimated by both approaches was also upwardly biased and the between-level component was also redistributed to the within-level. Additionally, *SEs* obtained by SLGCM and D-LGCM were also considerably biased and the latter approach tended to generate more biased *SEs*.

Parameter estimates and *SEs* in the between-level covariance structure

Parameters in the between-level covariance structure (γ_{00} , γ_{11} , and γ_{10}) could be only estimated by the MLGCM, and hence these three parameters were not the focus of the study. We present the parameter estimates and *SEs* of these parameters as well as corresponding relative biases in Table 3. Results can inform readers regarding the performance of MLGCM in the associated different sample size conditions. In our study, γ_{00} denotes variability of initial mathematics literacy scores across cluster units (e.g., school), γ_{11} signifies variability of linear trajectories across cluster units, and γ_{10} represents the relationship between cluster-level initial mathematics literacy score and cluster-level linear trajectory. The estimates for the intercept factor variance (γ_{00} ; ranging from 27.64 to 28.79), slope factor variance (γ_{11} ; ranging from 2.69 to 2.80), and covariance between intercept and slope (γ_{10} ; ranging from 2.49 to 2.69) were fairly accurate across all sample size scenarios. Similarly, we found *SEs* across all sample size scenarios to be reasonable accurate.

Discussion and conclusion

The current study investigated the effectiveness of two model-based approaches (MLGCM and MM) and one design-based approach (D-LGCM) that had the potential to accommodate individual dependency in longitudinal data. The accuracy of parameter estimates and *SEs* related to intraindividual growth (α_1 and α_2) and interindividual variability (π_{00} , π_{11} , and π_{10}) was evaluated through their corresponding relative bias. Moreover, the model solutions of SLGCM were also presented to acknowledge the consequences of ignoring individual dependency.

Intraindividual growth

In the current study, intraindividual growth can be indicated by two parameters, namely α_1 and α_2 , in the mean structure. As previously mentioned, parameters in the mean structure were estimated in the between-level model by the MLGCM and estimated at the within-level model by MM and D-LGCM. The results showed MLGCM can produce accurate estimated α_1 and α_2 as well as their *SEs* in all simulation scenarios. This finding was in line with Wu et al.'s (2015) and Muthén's (1997) studies, which also suggested

MLGCM was able to estimate intraindividual growth parameters accurately. In terms of the sample size required for accurate estimates, our study suggested MLGCM can perform satisfactorily even in a small NC (30) and CS (5) condition. Similarly, estimated α_1 and α_2 and corresponding *SEs* derived from MM and D-LGCM can also be considered trustworthy. Thus, when intraindividual growth is of interest, researchers are free to implement either MM or D-LGCM on clustered longitudinal data.

On the contrary, we found that although SLGCM provided unbiased estimates of α_1 and α_2 , the *SEs* were seriously underestimated. Our findings on SLGCM were consistent with prior reports in Muthén's (1997) paper. Note the downward bias of *SE* can cause larger Type I error rates for researchers attempting to conduct inferential tests. As a result, when studying average initial score and average linear trajectory is of primary interest, we recommend researchers consider multilevel approaches, either model-based or design-based ones, that can take into account the individual dependency of the longitudinal data.

Interindividual variability

Parameter estimates of π_{00} , π_{11} , and π_{10} and their *SEs* in the individual-level covariance structure were used to determine interindividual variability. We found MLGCM satisfactorily provided accurate parameter estimates and *SEs* for π_{00} and π_{11} regardless of the design factors, NC, and CS. Therefore, we suggest MLGCM can be used to analyze interindividual variability in initial scores and in linear trajectories. Furthermore, our results showed MLGCM tended to overestimate π_{10} (relationship between initial score and trajectory), but its *SEs* were not biased. Note π_{10} was set to be near-zero (0.32) in the population model (i.e., it can be converted into a correlation equal to .01, a negligible effect size). We did find all estimates of π_{10} were not statistically significant at $\alpha = .05$ level across all simulated replications and ANOVA results failed to support that estimates of π_{10} were associated with NC and CS, suggesting no systematic pattern. We recommend that the upwardly biased estimates of π_{10} be re-examined in future studies.

MM performed quite similarly as MLGCM did except for the parameter estimates of π_{10} . This approach in general provided unbiased parameter estimates and *SEs* of π_{00} and π_{11} . Unlike MLGCM, MM was more likely to underestimate π_{10} , but the corresponding *SEs* were still unbiased. As previously mentioned, the effect size of π_{10} in the population model was extremely small and we did not find any statistically significant estimates of π_{10} in our simulated replications. Moreover, ANOVA results showed estimates of π_{10} were not accounted for by NC and CS. Future studies can further examine the downward biased estimates of π_{10} in MM. Generally speaking, if researchers are interested in studying interindividual

variability in initial scores and in linear trajectories, both MLGCM and MM were capable of producing accurate parameter estimates and *SEs* and thus can be adopted for such analyses.

It must be noted that when D-LGCM and SLGCM were used to study interindividual variability, parameter estimates of π_{00} , π_{11} , and π_{10} did not consist exclusively of the variability in individual (e.g., students) growth but instead were comprised of the integrated variability of individual and cluster (e.g., school) growth, which cannot be interpreted. In line with Muthén (1997), we found ignoring the individual dependency, the factor variance (π_{00} and π_{11}) or covariance (π_{10}) belonging to the between-level was redistributed to the corresponding portions of the within-level model. Therefore, unless variability across cluster units in the between-level model (i.e., γ_{00} , γ_{11} , and γ_{10}) is close or equal to zero in the population, SLGCM and D-LGCM can reduce the accuracy of parameter estimates of π_{00} , π_{11} , and π_{10} to some degree. More specifically, because intercept variances (π_{00} and γ_{00}) and slope variances (π_{11} and γ_{11}) at both levels cannot be negative, the π_{00} (and π_{11}) at the within-level model estimated by D-LGCM and SLGCM will be upwardly biased unless the ignored γ_{00} (and γ_{11}) at the between-level of the population model are close or equal to zero.

On the other hand, π_{10} at the within level can be either upwardly or downwardly biased because the ignored γ_{10} at the between level of the population model can be either positive or negative. The biased estimates of π_{10} can lead to a misunderstanding of the relationship between individual initial scores and growth trajectories, which can be harmful to substantive research. For example, the idea of the Matthew effect in reading has gained considerable attention in the past few decades (Pfof, Hattie, Dörfler, & Artelt, 2014). The Matthew effect in reading is a positive correlation between initial competence level and individual development, which can be statistically estimated by the parameter π_{10} . In other words, an accurate estimate of π_{10} is essential to confirm the presence of the Matthew effect. Pfof et al. (2014) reviewed 28 studies focusing on the Matthew effect in reading in a meta-analysis. We reinvestigated the same studies and found some substantive researchers were not aware that SLGCM and D-LGCM could produce seriously biased π_{10} when clustered longitudinal data were analyzed. More specifically, we found clustered longitudinal data was used to assess the Matthew effect in 22 studies. However, 16 of 22 studies failed to take into account the clustering effect (e.g., using SLGCM); and two of 22 applied D-LGCM to analyze the data. Only four studies took into account the individual dependency in the data (three used MLGCM; one applied hierarchical linear modeling approach, which was beyond the scope of our study). Based on our simulation finding that π_{10} can be either upwardly or downwardly biased, we raise our concern that those 18 studies might have incorrectly determined the presence or absence of

the Matthew effect in reading. Future substantive studies that attempt to understand the relationship between individual initial scores and growth trajectories using clustered longitudinal data but fail to account for the structure appropriately need to be aware of the potential bias introduced.

We further explored whether the chi-square (χ^2) statistic could effectively detect models with biased parameter estimates of π_{00} , π_{11} , and π_{10} in our simulated replications of either D-LGCM or SLGCM. Unfortunately, the results showed only 4.64% and 4.68% of replications for SLGCM and D-LGCM, respectively, were rejected by the χ^2 statistic with α at .05 level. Practically, researchers could be misled by the χ^2 statistic. The take-home message is that D-LGCM does not qualify as an alternative approach to analyze clustered longitudinal data if interindividual variability is of interest.

What if either D-LGCM or SLGCM was used to analyze clustered longitudinal data? Practically, parameter estimates of interindividual variability often drives researchers to investigate further whether this variability can be explained by individual-level covariates (e.g., gender). Analytically, researchers can regress interindividual variability (e.g., π_{00} and π_{11}) on selected individual-level predictors. A previous study showed regression weights would not necessarily be biased (Wu et al., 2015); however, the explained variance of π_{00} and π_{11} accounted for by the individual-level predictors (Raudenbush & Bryk, 2002) could be misleading. Accordingly, using D-LGCM and SLGCM to analyze clustered longitudinal data is not appropriate and should be avoided.

An empirical example

An empirical example is useful in illustrating that MLGCM and MM are able to produce accurate parameter estimates and *SEs* of interindividual variability, while D-LGCM and SLGCM are not. This example analyzed data drawn from the Longitudinal Study of American Youth (LSAY; Miller et al., 2000), which is a national panel study of mathematics and science education in US public schools. The LSAY has been widely used to study the growth of mathematics and science performance (e.g., Ma & Ma, 2004; Ma & Wilkins, 2007). Following Muthén's (2004) paper, we analyzed Cohort 2 data containing 3,102 students nested within 52 schools. Students' grade 7 to grade 10 mathematics achievement scores obtained by item response theory equating were analyzed by MLGCM (see Fig. 1), MM (see Fig. 2), D-LGCM, and SLGCM, separately. The parameter estimates are presented in Table 4. The *Mplus* syntax of four models is provided in Table 7 of Appendix.

The fit indices of the MLGCM [$\chi^2(8) = 119.957$, $p < .05$; Root Mean Square Error of Approximation (*RMSEA*) = .067, Comparative Fit Index (*CFI*) = .984, Tucker-Lewis Index (*TLI*) = .976; Standardized Root Mean Square Residual for within (*SRMR-W*) = .007; *SRMR* for between (*SRMR-B*) = .004] suggested adequate model fit (Hu & Bentler, 1999). Therefore, the

parameter estimates provided by the MLGCM were expected to be less biased. On the other hand, the fit indices of the MM also suggested adequate model fit [$\chi^2(5) = 109.873$, $p < .05$; $RMSEA = .082$, $CFI = .985$, $TLI = .963$; $SRMR-W = .007$; $SRMR-B = .003$]. Consistent with our simulation findings, the parameter estimates in the mean structure and within-covariance structure produced by MM were similar to those derived by MLGCM. Thus, both MLGCM and MM can be used to investigate intraindividual growth and interindividual variability given clustered longitudinal data.

D-LGCM produced parameter estimates in the mean structure similar to those derived by MLGCM. However, in line with our simulation findings, when D-LGCM was applied, the parameter estimates in within-covariance structure were greatly biased because factor (co)variance at the between-level was redistributed to the within-level. For example, the estimated within-level intercept factor variance in D-LGCM (87.37) was approximately equal to the sum of the within-level intercept factor variance (70.57) and between-level intercept factor variance (17.31) in MLGCM. Therefore, researchers should avoid using D-LGCM to study interindividual variability given clustered longitudinal data. When SLGCM was applied, the *SEs* in the mean structure were underestimated compared with those derived by MLGCM. Similar to D-LGCM, the parameter estimates in the within-covariance structure were seriously biased. Therefore, using D-LGCM to analyze clustered longitudinal data is not appropriate.

Limitations and future research direction

This study has limitations that should be addressed. First, the current study adopted a two-level population model depicting

the linear growth trajectory at the between- and within-level intercept factor variance. Therefore, findings should only be generalized to studies that apply similar models. Further studies are needed to determine whether the current findings can also be replicated using different models (e.g., a two-level quadratic growth model). Second, we considered a limited number of design factors in the present study. Additional scenarios created by using different design factors, such as unbalanced designs (unequal group conditions) and varying the number of observed indicators per latent factor, are needed in future studies. Third, the information regarding minimum sample size needed for the between-level and within-level intercept factor variance is quite useful for practitioners who attempt to apply a MLGCM to clustered longitudinal data. To the best of our knowledge, no study has been conducted on this issue. Based on our simulation findings, we found a minimum sample size ($NC = 30$ and $CS = 5$) equal to 150 is sufficient to attain precise parameter estimates and *SEs* in both within- and between-covariance structures and between-mean structure if the two-level model is correctly specified. Future studies are needed to investigate systematically this sample size topic in MLGCM. Last but not least, the parameter settings in the present simulation study were drawn from an empirical study (Baumert et al., 2012) to increase the generalizability of our findings. Therefore, the variability across cluster units at the between-level (i.e., γ_{00} , γ_{11} , and γ_{10}) intercept factor variance was fixed to one condition and was not manipulated. Future studies can consider the magnitude of variability across cluster units as a design factor and systematically investigate the acceptable degree of bias of interindividual variability (i.e., π_{00} , π_{11} , and π_{10}) derived by model-based and design-based approaches.

Appendix

Table 5 *Mplus* source code for the simulation study

```
TITLE: Data Generation using a Multilevel latent growth curve model (NC = 50 and CS = 5)
montecarlo:
names are y1-y4;
nobs = 250;
ncsizes = 1;
csizes = 50(5);
seed = 67584;
nreps = 1000;
repsave = all;
save = d:\50_5rep*.dat;
ANALYSIS:
TYPE IS TWOLEVEL;
MODEL POPULATION:
%WITHIN%
iw sw | y1@0 y2@1 y3@2 y4@3;
y1-y4*40.57;
iw@113.3;
sw@8.77;
```


Table 5 (continued)

```

iw with sw @0.315221;
%BETWEEN%
ib sb | y1@0 y2@1 y3@2 y4@3;
y1-y4@0;
[ib@96.28 sb@9.94];
ib@28.86;
sb@2.80;
ib with sb @2.606905;
MODEL:
%WITHIN%
iw sw | y1@0 y2@1 y3@2 y4@3;
y1-y4*40.57;
iw@113.3;
sw@8.77;
iw with sw @0.315221;
%BETWEEN%
ib sb | y1@0 y2@1 y3@2 y4@3;
y1-y4@0;
[ib@96.28 sb@9.94];
ib@28.86;
sb@2.80;
ib with sb @2.606905;

TITLE: Data Analysis using a Multilevel latent growth curve model (NC = 50 and CS = 5)
DATA: FILE IS d:\50_5rep1.dat;
VARIABLE:
NAMES ARE y1-y4 clus;
CLUSTER = clus;
ANALYSIS:
TYPE IS TWOLEVEL;
ESTIMATOR = MLR;
MITERATIONS =1000;
MODEL:
%WITHIN%
iw sw | y1@0 y2@1 y3@2 y4@3;
y1-y4*40.57;
iw*113.3;
sw*8.77;
iw with sw *0.315221;
%BETWEEN%
ib sb | y1@0 y2@1 y3@2 y4@3;
y1-y4@0;
[ib*96.28 sb*9.94];
ib*28.86;
sb*2.80;
ib with sb *2.606905;

TITLE: Data Analysis using a Maximum model (NC = 50 and CS = 5)
DATA: FILE IS d:\50_5rep1.dat;
VARIABLE:
NAMES ARE y1-y4 clus;
CLUSTER = clus;
ANALYSIS:
TYPE IS TWOLEVEL;
ESTIMATOR = MLR;
MITERATIONS =1000;
MODEL:
%WITHIN%
iw sw | y1@0 y2@1 y3@2 y4@3;
y1-y4*40.57;
iw*113.3;
sw*8.77;
iw with sw *0.315221;
[ib*96.28 sw*9.94];
%BETWEEN%
y1-y4 with y1-y4;

```

Table 5 (continued)

TITLE: Data Analysis using a Design-based latent growth curve model (NC = 50 and CS = 5)
 DATA: FILE IS d:\50_5rep1.dat;
 VARIABLE:
 NAMES ARE y1-y4 clus;
 CLUSTER = clus;
 ANALYSIS:
 TYPE = COMPLEX;
 MITERATIONS =1000;
 MODEL:
 iw sw | y1@0 y2@1 y3@2 y4@3;
 y1-y4*40.57;
 [iw*96.28 sw*9.94];
 iw*113.3;
 sw*8.77;
 iw with sw *0.315221;
 TITLE: Data Analysis using a Single-level latent growth curve model (NC = 50 and CS = 5)
 DATA: FILE IS d:\50_5rep1.dat;
 VARIABLE:
 NAMES ARE y1-y4 clus;
 USEVARIABLE = y1-y4;
 ANALYSIS:
 MITERATIONS =1000;
 MODEL:
 iw sw | y1@0 y2@1 y3@2 y4@3;
 y1-y4*40.57;
 [iw*96.28 sw*9.94];
 iw*113.3;
 sw*8.77;
 iw with sw *0.315221;

Table 6 Empirical power of key parameters in the population model

		Mean structure		Covariance structure of within-level model			Covariance structure of between-level model		
		α_1 (mean of intercept factor)	α_2 (mean of slope factor)	π_{00} (variance of intercept factor)	π_{11} (variance of slope factor)	π_{10} (covariance between two factors)	γ_{00} (variance of intercept factor)	γ_{11} (variance of slope factor)	γ_{10} (covariance between two factors)
<i>Population value</i>		<i>96.28</i>	<i>9.94</i>	<i>113.30</i>	<i>8.77</i>	<i>0.32</i>	<i>28.86</i>	<i>2.80</i>	<i>2.61</i>
NC	CS								
30	5	1.000	1.000	1.000	0.952	0.080	0.469	0.364	0.088
30	10	1.000	1.000	1.000	1.000	0.078	0.881	0.778	0.197
30	20	1.000	1.000	1.000	1.000	0.080	0.986	0.974	0.247
50	5	1.000	1.000	1.000	0.996	0.057	0.769	0.649	0.157
50	10	1.000	1.000	1.000	1.000	0.066	0.990	0.962	0.291
50	20	1.000	1.000	1.000	1.000	0.080	1.000	1.000	0.354
100	5	1.000	1.000	1.000	1.000	0.057	0.980	0.944	0.295
100	10	1.000	1.000	1.000	1.000	0.056	1.000	1.000	0.477
100	20	1.000	1.000	1.000	1.000	0.080	1.000	1.000	0.613
150	5	1.000	1.000	1.000	1.000	0.058	0.999	0.995	0.403
150	10	1.000	1.000	1.000	1.000	0.064	1.000	1.000	0.643
150	20	1.000	1.000	1.000	1.000	0.087	1.000	1.000	0.794
200	5	1.000	1.000	1.000	1.000	0.061	1.000	1.000	0.503

Table 6 (continued)

		Mean structure		Covariance structure of within-level model			Covariance structure of between-level model		
		α_1 (mean of intercept factor)	α_2 (mean of slope factor)	π_{00} (variance of intercept factor)	π_{11} (variance of slope factor)	π_{10} (covariance between two factors)	γ_{00} (variance of intercept factor)	γ_{11} (variance of slope factor)	γ_{10} (covariance between two factors)
200	10	1.000	1.000	1.000	1.000	0.070	1.000	1.000	0.769
200	20	1.000	1.000	1.000	1.000	0.097	1.000	1.000	0.893
500	5	1.000	1.000	1.000	1.000	0.069	1.000	1.000	0.848
500	10	1.000	1.000	1.000	1.000	0.091	1.000	1.000	0.987
500	20	1.000	1.000	1.000	1.000	0.106	1.000	1.000	1.000

Table 7 *Mplus* source code for the empirical example

```

TITLE: Data Analysis using a Multilevel latent growth curve model
DATA: FILE IS LASY_mplus.dat;
VARIABLE:
NAMES ARE stuid schid m7 m8 m9 m10;
USEVARIABLES ARE m7-m10;
CLUSTER = schid;
ANALYSIS:
TYPE IS TWOLEVEL;
ESTIMATOR = MLR;
MITERATIONS =1000;
MODEL:
%WITHIN%
iw sw| m7@0 m8@1 m9@2 m10@3;
%BETWEEN%
ib sb| m7@0 m8@1 m9@2 m10@3;
m7-m10;
[ib sb];
OUTPUT:
tech8;

TITLE: Data Analysis using a Maximum model
DATA: FILE IS LASY_mplus.dat;
VARIABLE:
NAMES ARE stuid schid m7 m8 m9 m10;
USEVARIABLES ARE m7-m10;
CLUSTER = schid;
ANALYSIS:
TYPE IS TWOLEVEL;
ESTIMATOR = MLR;
MITERATIONS =1000;
MODEL:
%WITHIN%
iw sw| m7@0 m8@1 m9@2 m10@3;
[iw sw];
%BETWEEN%
m7-m10 with m7-m10;
OUTPUT:
tech8;

TITLE: Data Analysis using a Design-based latent growth curve model
DATA: FILE IS LASY_mplus.dat;
VARIABLE:
NAMES ARE stuid schid m7 m8 m9 m10;
USEVARIABLES ARE m7-m10;
CLUSTER = schid;
ANALYSIS:

```

Table 7 (continued)

TYPE = COMPLEX;
 ESTIMATOR = MLR;
 MITERATIONS =1000;
 MODEL:
 i s| m7@0 m8@1 m9@2 m10@3;
 TITLE: Data Analysis using a Single-level latent growth curve model
 DATA: FILE IS LASY_mplus.dat;
 VARIABLE:
 NAMES ARE stuid schid m7 m8 m9 m10;
 USEVARIABLES ARE m7-m10;
 ESTIMATOR = MLR;
 MITERATIONS =1000;
 MODEL:
 i s| m7@0 m8@1 m9@2 m10@3;

References

- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148.
- Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development*, 83(4), 1347–1367. doi:10.1111/j.1467-8624.2012.01779.x
- 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*. Mahwah: Lawrence Erlbaum Associates.
- Chen, Q., Kwok, O.-M., Luo, W., & Willson, V. L. (2010). The impact of ignoring a level of nesting structure in multilevel growth mixture models: A Monte Carlo study. *Structural Equation Modeling*, 17(4), 570–589. doi:10.1080/10705511.2010.510046
- Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, 11(2), 121–136. doi:10.1080/15248371003699969
- Diallo, T. M. O., Morin, A. J. S., & Parker, P. D. (2014). Statistical power of latent growth curve models to detect quadratic growth. *Behavior Research Methods*, 46(2), 357–371. doi:10.3758/s13428-013-0395-1
- Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller, M., & Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research*, 32(3), 275–318. doi:10.1207/s15327906mbr3203_3
- Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An Introduction to Latent Variable Growth Curve Modeling Concepts, Issues, and Applications* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.
- Ferragut, M., Blanca, M. J., & Ortiz-Tallo, M. (2014). Psychological virtues during adolescence: A longitudinal study of gender differences. *European Journal of Developmental Psychology*, 11(5), 521–531. doi:10.1080/17405629.2013.876403
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. New York: Guilford Press.
- Heck, R. H., & Thomas, S. L. (2009). *An introduction to multilevel modeling techniques* (2nd ed.). New York: Routledge.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied survey data analysis*. Boca Raton: Chapman & Hall/CRC.
- Hertzog, C., Oertzen, T. v., & Ghisletta, P. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling*, 15, 541–563. doi:10.1080/10705510802338983
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329–367. doi:10.1177/0049124198026003003
- Hox, J. J. (2010). *Multilevel Analysis Techniques and Applications* (2nd ed.). New York: Routledge.
- Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, 8(2), 157–174. doi:10.1207/S15328007SEM0802_1
- 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.
- Ingels, S. J., Pratt, D. J., Rogers, J. E., Siegel, P. H., & Stutts, E. S. (2005). *Education Longitudinal Study of 2002: Base-year to first follow-up data file documentation (NCES 2006–344)*. Retrieved from Washington, DC.
- Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, 8(3), 325–352. doi:10.1207/S15328007SEM0803_1
- Kwok, O., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research*, 42(3), 557–592. doi:10.1080/00273170701540537
- Lai, M. H. C., & Kwok, O. (2015). Examining the rule of thumb of not using multilevel modeling: The “design effect smaller than two” rule. *The Journal of Experimental Education*, 83(3), 423–438. doi:10.1080/00220973.2014.907229
- 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
- Ma, X., & Ma, L. (2004). Modeling stability of growth between mathematics and science achievement during middle and high school. *Evaluation Review*, 28(2), 104–122. doi:10.1177/0193841X03261025
- Ma, X., & Wilkins, J. M. (2007). Mathematics coursework regulates growth in mathematics achievement. *Journal for Research in Mathematics Education*, 38(3), 230–257.
- Meyers, J. L., & Beretvas, N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, 41(4), 473–497. doi:10.1207/s15327906mbr4104_3
- Miller, J. D., Kimmel, L., Hoffer, T. B., & Nelson, C. (2000). *Longitudinal study of American youth: User's manual*. Evanston: Northwestern University, International Center for the Advancement of Scientific Literacy.

- Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, 39(1), 129–149. doi:[10.1207/s15327906mbr3901_5](https://doi.org/10.1207/s15327906mbr3901_5)
- Muthén, B. O. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (Ed.), *Latent variable modeling and applications to causality* (pp. 149–161). New York, NY.
- Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 345–368). Newbury Park: Sage Publications.
- Muthén, B. O., & Asparouhov, T. (2009). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis*. New York: Routledge.
- Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *The Handbook of Advanced Multilevel Analysis*. New York: Routledge.
- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2(4), 371–402. doi:[10.1037/1082-989X.2.4.371](https://doi.org/10.1037/1082-989X.2.4.371)
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267–316. doi:[10.2307/271070](https://doi.org/10.2307/271070)
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guild* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599–620.
- Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew Effects in reading. *Review of Educational Research*, 84(2), 203–244. doi:[10.3102/0034654313509492](https://doi.org/10.3102/0034654313509492)
- Pomprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: Some consequences for model fit and standardized. *Multivariate Behavioral Research*, 49, 518–543. doi:[10.1080/00273171.2014.933762](https://doi.org/10.1080/00273171.2014.933762)
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks: Sage.
- Schaie, K. W. (1983). *Longitudinal studies of adult psychological development*. New York: Guilford Press.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis, modeling change and event occurrence*. New York: Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks: Sage.
- 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: Information Age.
- Stapleton, L. M. (2008). Variance estimation using replication methods in structural equation modeling with complex sample data. *Structural Equation Modeling*, 15(2), 183–210. doi:[10.1080/10705510801922316](https://doi.org/10.1080/10705510801922316)
- Wu, J.-Y., & Kwok, O. (2012). Using SEM to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. *Structural Equation Modeling*, 19(1), 16–35. doi:[10.1080/10705511.2012.634703](https://doi.org/10.1080/10705511.2012.634703)
- Wu, J.-Y., Kwok, O., & Willson, V. L. (2015). Using design-based latent growth curve modeling with cluster-level predictor to address dependency. *Journal of Experimental Education*, 82(4), 431–454. doi:[10.1080/00220973.2013.876226](https://doi.org/10.1080/00220973.2013.876226)
- Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior Research Methods*, 41(4), 1083–1094. doi:[10.3758/BRM.41.4.1083](https://doi.org/10.3758/BRM.41.4.1083)