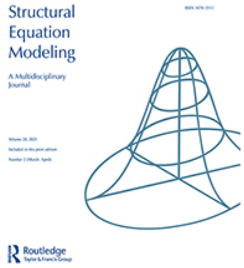

Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

A Cautionary Note on Identification and Scaling Issues in Second-order Latent Growth Models

Yanyun Yang, Yachen Luo & Qian Zhang

To cite this article: Yanyun Yang, Yachen Luo & Qian Zhang (2021) A Cautionary Note on Identification and Scaling Issues in Second-order Latent Growth Models, Structural Equation Modeling: A Multidisciplinary Journal, 28:2, 302-313, DOI: 10.1080/10705511.2020.1747938

To link to this article: https://doi.org/10.1080/10705511.2020.1747938

Published online: 05 Jun 2020.

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
2021, VOL. 28, NO. 2, 302–313
https://doi.org/10.1080/10705511.2020.1747938

A Cautionary Note on Identification and Scaling Issues in Second-order Latent Growth Models

Yanyun Yang, Yachen Luo, and Qian Zhang

Florida State University

ABSTRACT
The second-order latent growth models (2nd-order LGMs) have been recommended to analyze longitudinal data when latent constructs are measured by multiple indicators. However, identification and scaling issues in 2nd-order LGMs have not been well understood. Using both formulas and a numerical example, we show that under strong longitudinal factor invariance estimates of growth parameters in 2nd-order LGMs depend on how the latent factor is scaled. With the marker-variable identification method, estimates of growth parameters depend on the choice of marker variable and the constants applied to its loading and intercept. With the effect-coding identification method, three sets of constraints applied to loadings and intercepts lead to meaningful interpretation of growth trajectories. When the latent factor at the reference time point is standardized, estimates of growth parameters are unique. We suggest users explicitly state the identification and scaling method in interpreting estimated growth parameters when a 2nd-order LGM is applied.

KEYWORDS
Second-order latent growth model; marker-variable identification; effect-coding identification; latent-standardization identification

Latent growth modeling is referred to as a class of statistical techniques for analyzing longitudinal data (Bollen & Curran, 2006). It allows researchers to examine how some aspects of individuals change over time, and how the change process may differ across individuals. For example, Knight et al. (2018) examined the development of familism values from 5th to 10th grade using a sample of Mexican American adolescents. When modeling longitudinal data similar to those in Knight et al. (2018), researchers are mostly interested in estimates of growth parameters, which capture the averaged intra-individual change and inter-individual difference in intra-individual changes of constructs (e.g., familism values) over time. Figure 1 depicts a latent growth model (LGM) in which a linear growth trajectory is imposed on a construct across four equally spaced measurement occasions (indicated by x_1-x_4) and the first measurement occasion is the reference time point. The five growth parameters of interest are: the mean of the level factor (κ_L), the mean of the slope factor (κ_S), and the variance and covariance of growth factors (ϕ_{LL}, ϕ_{SS}, and ϕ_{LS}). These symbols will be further defined in the next section.

The latent growth model shown in Figure 1 is referred to as the first-order latent growth model (1st-order LGM) because an *observed* variable is used to represent the construct of interest on which the growth trajectory is examined. In psychology and education disciplines, many constructs are not directly observable; instead, they are measured by multiple items which are conceptualized as reflective indicators of the constructs. A common practice in longitudinal analysis is to compute a composite variable as the mean or the sum of item scores and then use the composite variable to represent the construct in LGMs (Leite, 2007). Using a composite variable as an approximation of the latent construct in an LGM is based on two assumptions. First, items are perfectly reliable because measurement errors are not taken into consideration in the modeling process. This assumption is untenable in most of psychology and educational research. Second, items measure the latent construct in the same way across measurement occasions, that is, longitudinal measurement invariance (MI; Chan, 1998) is assumed. This assumption cannot be tested in a 1st-order LGM because the measurement component of the latent construct is not included in the model.

The longitudinal MI can be tested in a second-order latent growth model (2nd-order LGM; Hancock et al., 2001). In addition, a 2nd-order LGM takes measurement errors into consideration and thus does not require items being perfectly reliable. In a 2nd-order LGM, the first-order component describes the relationship between items and latent factors, mainly dealing with measurement issues (e.g., reliability, longitudinal measurement invariance). The second-order component allows for testing a hypothesized growth trajectory by imposing constraints to the covariance and mean structures of latent factors over time. McArdle (1988) named this type of LGMs the curve-of-factors model in the sense that the growth curve is evaluated on the basis of latent factors. An example 2nd-order LGM is shown in Figure 2 and is used throughout this paper. Different from the model in Figure 1, in the 2nd-order LGM the latent construct is measured by four items and its relationship with items is explicitly modeled.

Compared to 1st-order LGMs, however, 2nd-order LGMs have been much less frequently applied in empirical studies, and issues of identification and scaling for latent variables in

CONTACT Yanyun Yang yyang3@fsu.edu Department of Educational Psychology & Learning Systems, Florida State University, Tallahassee, FL 32306.

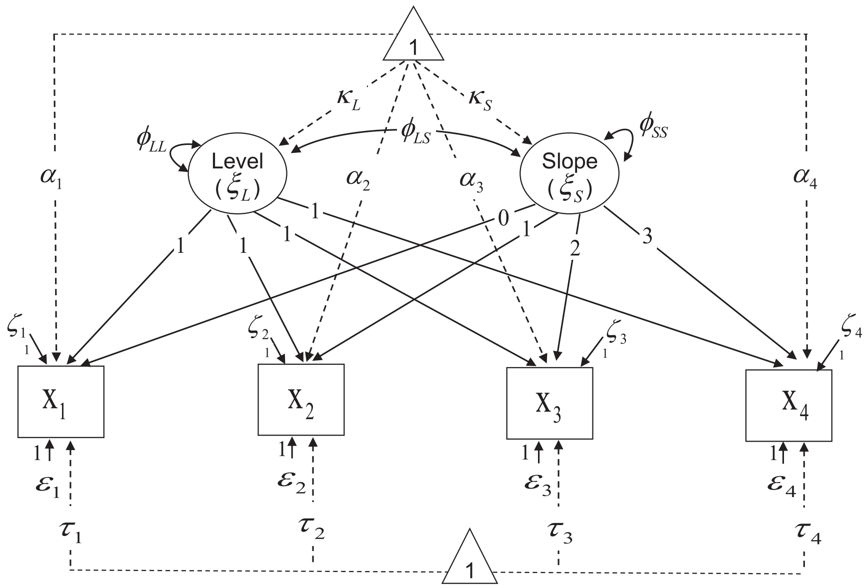

Figure 1. An example of the first-order latent growth model.
Model parameters related to the covariance structure are indicated by the solid lines; model parameters related to the mean structure are indicated by the dash lines.

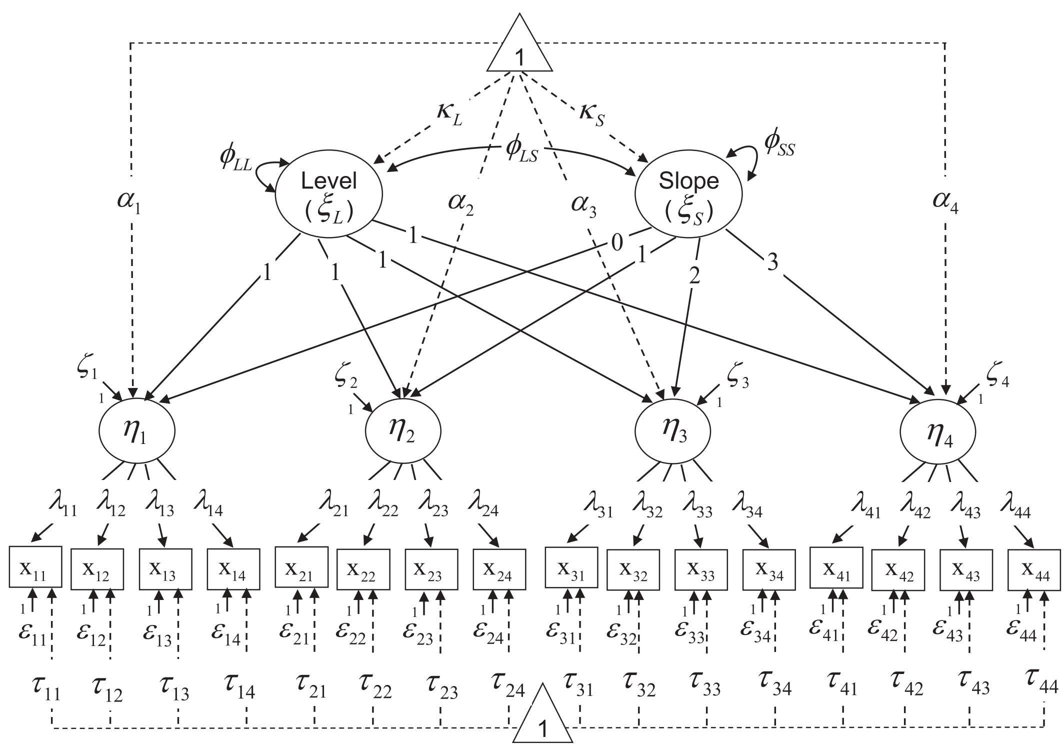

Figure 2. An example of the second-order latent growth model.
Model parameters related to the covariance structure are indicated by the solid lines; model parameters related to the mean structure are indicated by the dash lines.

this type of models have not been well understood by applied researchers. Understanding how latent variables are identified and assigned a scale and how different identification methods impact the growth parameter estimation and interpretation is important. This is because growth parameters in a 2^{nd}-order LGM are estimated on the basis of latent factors. Each latent factor has to be assigned a scale (e.g., mean and variance) in an analysis. Potentially, there are infinite ways to scale latent factors, given that the model is identified. Different approaches to assigning each latent factor a scale give different sets of unstandardized growth parameter estimates, which are often reported in empirical studies.

To get a picture of how 2^{nd}-order LGMs have been applied in empirical studies, we conducted a review in four psychological journals: *Journal of Personality and Social Psychology, Journal of Applied Psychology, Developmental Psychology*, and *Child Development*. Searching with the keywords "growth curve model" and "latent growth model" from the first issue to November 2019, we located a total of 334 papers. Thirty-four of the papers were excluded because they did not use LGM in analyses. Among the 300 papers that applied LGMs, only 11 papers used 2^{nd}-order LGMs. Three of the 11 papers showed in path diagrams that latent factors were assigned a scale by fixing the loading of the first measurement indicator to be one (i.e., applying the marker-variable identification method detailed later). The other eight papers did not explicitly state or show which method was used to assign a scale for latent factors, but seven of them mentioned that M*plus* was used for analyses. We suspected that these seven papers also used the marker-variable identification method with the first indicator as the marker variable because this method is implemented as the default setting in M*plus*.

Purpose of the current study

Identification and scaling methods have been well discussed in structural equation modeling, in particular, factor analysis and multiple-group MI analysis (e.g., Klößner & Klopp, 2019; Little et al., 2007), but have not been well elaborated in 2^{nd}-order LGMs. Based on our review, it appears that the marker-variable identification approach with the first indicator as the marker variable was most frequently used. However, growth parameter estimates and the consequent interpretation of growth trajectory depend on the choice of the marker variable. For example, of the three papers that explicitly showed the first indicator being the marker variable, Knight et al. (2018) examined the development of familism values across 5^{th}, 7^{th}, and 10^{th} grades with 749 Mexican American adolescents. Familism values were measured by four indicators: support (the first indicator), obligations, family as referent, and respect. Each indicator was scored in the range from 1 to 5. The authors reported that "the second-order linear growth model closely fit the data … Adolescents varied significantly in their fifth-grade familism values (intercepts) … and change in familism values from fifth to 10^{th} grade (slopes). The mean familism score at the fifth grade was quite high (4.5 on a 5-point scale) … and the mean slope indicated a decrease in familism values of. 04 points per year." (p. 381). Similar reporting practice and interpretation can be found in other applications of 2^{nd}-order LGMs (e.g., Robinson et al., 2018; Weidinger et al., 2018).

Had the authors chosen another indicator as the marker variable, growth parameter estimates and consequently interpretation of the growth trajectory would have been different. In the paper, we use both formulas and an illustrative example to show under longitudinal MI,[1] how different identification and scaling methods may yield different estimates of growth parameters and interpretations of growth trajectories. We focus on three approaches to identifying and scaling latent factors in 2^{nd}-order LGMs: the marker-variable identification method, the effect-coding identification method, and the latent-standardization identification method. We hope that this paper may (1) increase applied researchers' awareness of the dependence of growth parameter estimates on identification and scaling methods, (2) help applied researchers understand that estimates and statistical testing of parameter estimates are not meaningful without being referenced to the scale of latent factors, and (3) avoid misinterpreting model results.

In what follows, we first present an example 2^{nd}-order LGM as shown in Figure 2. We then discuss three methods of scaling latent factors,[2] and show how the estimates and interpretation of growth parameters depend on the scaling of latent factors. We conclude the paper with some suggestions with respect to applications of 2^{nd}-order LGMs.

Second-order latent growth models

A 2^{nd}-order LGM is a constrained 2^{nd}-order factor analysis model. Under strong longitudinal MI, the first-order component of the model shown in Figure 2 is expressed as:

$$\begin{bmatrix} x_{11} \\ \cdots \\ x_{14} \\ x_{21} \\ \cdots \\ x_{24} \\ x_{31} \\ \cdots \\ x_{34} \\ x_{41} \\ \cdots \\ x_{44} \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \cdots \\ \tau_4 \\ \tau_1 \\ \cdots \\ \tau_4 \\ \tau_1 \\ \cdots \\ \tau_4 \\ \tau_1 \\ \cdots \\ \tau_4 \end{bmatrix} + \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \cdots & 0 & 0 & 0 \\ \lambda_4 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & \cdots & 0 & 0 \\ 0 & \lambda_4 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & \lambda_1 \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} * \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \cdots \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \cdots \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \cdots \\ \varepsilon_{34} \\ \varepsilon_{41} \\ \cdots \\ \varepsilon_{44} \end{bmatrix} \quad (1)$$

and the second-order component of the model is expressed as:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} * \begin{bmatrix} \xi_L \\ \xi_S \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix}, \quad (2)$$

where x_{tj} indicates the score on item j (j = 1, 2, 3, 4) at time t (t = 1, 2, 3, 4), τ is item intercept, λ is factor loading, η is factor score, α denotes factor intercept, ξ is growth factor, ε is measurement error, and ζ indicates the factor uniqueness. Both ε and ζ are assumed to have mean of zero and are uncorrelated with η and ξ.

[1]Only when the longitudinal MI holds can changes in observed item scores be attributed to changes in latent factors instead of changes in measurement properties over time. Longitudinal MI is a prerequisite of investigating the growth trajectory of a latent construct over time. In our derivation and illustration, we assume that the strong factorial invariance (i.e., each item has equal loading and equal intercept over time) is met.

[2]Item uniqueness and factor uniqueness are identified by fixing the path coefficient from uniqueness to item or factor to one, which is the default setting in M*plus*.

The expectation of item scores and covariance matrix among items are

$$\begin{aligned}\mathbf{u}_{xx} &= \tau + \boldsymbol{\Lambda}\mathrm{E}(\eta) = \tau + \boldsymbol{\Lambda}(\alpha + \mathbf{B}\boldsymbol{\kappa}) \ , \\ \boldsymbol{\Sigma}_{xx} &= \boldsymbol{\Lambda}(\mathbf{B}\Phi\mathbf{B}' + \boldsymbol{\Theta})\boldsymbol{\Lambda}' + \boldsymbol{\Psi}\end{aligned} \tag{3}$$

where τ is a 16×1 item intercept vector, $\boldsymbol{\Lambda}$ is a 16×4 factor loading matrix, α is a 4×1 factor intercept vector and is fixed as a zero vector to render a meaningful interpretation of growth parameters, $\mathbf{B}$ is a 4×2 matrix containing pattern coefficients characterizing the linear growth trajectory, $\boldsymbol{\kappa}$ is a 2×1 mean vector of growth factors, $\boldsymbol{\Phi}$ is a 2×2 covariance matrix among growth factors, $\boldsymbol{\Theta}$ is a 4×4 covariance matrix among factor uniqueness, and $\boldsymbol{\Psi}$ is a 16×16 covariance matrix among measurement errors. $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ are diagonal matrices.[3]

Both mean and covariance structures are involved in a 2nd-order LGM. Model parameters in the mean structure include elements in τ and $\boldsymbol{\kappa}$. Model parameters in the covariance structure include elements in $\boldsymbol{\Lambda}$, $\mathbf{B}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Psi}$. The 2nd-order LGM is not identified without placing restrictions on parameters for two reasons. First, there are many more parameters than the number of unique information that the data can offer, that is, the rule of $df \geq 0$ is not satisfied. Second, latent variables are not identified and not interpretable without being assigned a scale. Latent variables consist of measurement errors, factors measured by measurement indicators, factor uniqueness, and growth factors. In the next section, we discuss three approaches to identifying and scaling latent factors using the example 2nd-order LGM shown in Figure 2. We focus on five growth parameters that are of interest in the second-order linear latent growth model: the mean of the level factor (κ_L), the mean of the slope factor (κ_S), the variance of growth factors (ϕ_{LL} and ϕ_{SS}), and the covariance of growth factors (ϕ_{LS}).

To further illustrate the dependency of growth parameter estimates on identification methods, we use a numerical example by selecting the following parameter values based on the model shown in Figure 2:

- The mean vector for the level and slope factors is: $\boldsymbol{\kappa}' = [\kappa_L \ \kappa_S] = [1 \ 1]$.
- The covariance matrix of the level and slope factors is: $\boldsymbol{\Phi} = \begin{bmatrix} \phi_{LL} & \\ \phi_{LS} & \phi_{SS} \end{bmatrix} = \begin{bmatrix} .5 & \\ .1 & .1 \end{bmatrix}$.
- The covariance matrix among factor uniqueness is a diagonal matrix with $diag(\boldsymbol{\Theta})$= [2.0, 2.2, 1.7, 1.5].
- The intercept of the latent factor at each time point is zero, that is, α= [0, 0, 0, 0].
- The pattern coefficient matrix characterizing a linear growth trajectory is $\mathbf{B}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$.
- The factor loadings of the four items are [.6, .6, .8, .8] and are invariant across time.
- The intercepts of the four items are [.5, .5, .7, .7] and are invariant across time.
- The residual covariance matrix among items is a diagonal matrix with $diag(\boldsymbol{\Psi})$= [.6, .6, .4, .4, .5, .5, .6, .6, .7, .7, .8, .8, .8, .8, .7, .7].

Based on Equation (3), these parameter values result in means and covariance matrix among items reported in Table 1. In the current paper, illustration is based on population means and covariance matrix. We are not interested in comparisons across different identification methods in finite samples; therefore, we do not randomly draw sample data with various sizes.

Based on the classical true score theory and Equation (1), the score variance for item j at time t ($\sigma^2_{x_{tj}}$) is the sum of the true score variance ($\lambda^2_{tj}\sigma^2_{\eta_t}$) and the error score variance (ψ_{tj}). Reliability of item scores is computed as:

$$\rho_{x_{tj}} = \frac{\sigma^2_{x_{tj}} - \psi_{tj}}{\sigma^2_{x_{tj}}} = \frac{\lambda^2_{tj}\sigma^2_{\eta_t}}{\sigma^2_{x_{tj}}} \ . \tag{4}$$

The last row in Table 1 reports reliability coefficients for individual items, ranging from .60 to .80. Such a variance decomposition also applies to composite scores and is elaborated under the section "The Effect-Coding Identification Method." We may also apply the procedure to latent factors and compute reliability of growth at each time point and reliability of entire longitudinal sequence.[4]

[3]Some error covariances in Ψ may be allowed in applications (e.g., nonzero covariances of measurement errors for the same item over time).

[4]Technical details for computing reliability of growth at each time point and of entire longitudinal sequence are presented in Willett (1989) and Laenen et al. (2009). We apply this procedure to 2nd-order LGM. Based on Equation (2), reliability of growth at time t is computed by:

$$\rho_{\eta_t} = \left(\sigma^2_{\eta_t} - \theta_t\right) \Big/ \sigma^2_{\eta_t},$$

where $\sigma^2_{\eta_t}$ and θ_t denote variance of latent factor and variance of factor uniqueness at time t, respectively. Applied to our numerical example, the reliability coefficients are .200, .267, .433, and .571 from the first time point to the fourth time point, respectively. Each value can be interpreted as the percentage of variance in the latent factor at time t that is explained by the growth factors. Reliability of entire longitudinal sequence is computed by:

$$\rho_{\eta\eta} = \frac{\mathrm{l}'(\mathbf{B}\Phi\mathbf{B}')\mathrm{l}}{\mathrm{l}'(\mathbf{B}\Phi\mathbf{B}' + \Theta)\mathrm{l}},$$

where l denotes a unit vector and the rest of the symbols are defined in Equation (3). Applied to our numerical example, the value is .689. This value can be interpreted as the percentage of total factor score variance across time that is explained by the growth factors. See Marcoulides (2019) for a working example for computing reliability of growth at each time point and of entire longitudinal sequence in the context of 1st-order LGM. As detailed in Willett (1989), reliability of entire longitudinal sequence tends to increase as the number of measurement occasions increases, variance of growth factors increases, and the variance of factor uniqueness decreases.

Table 1. Mean, covariance matrix, and reliability of the four items repeatedly measured over four time points.

	x_{11} 1	x_{12} 2	x_{13} 3	x_{14} 4	x_{21} 5	x_{22} 6	x_{23} 7	x_{24} 8	x_{31} 9	x_{32} 10	x_{33} 11	x_{34} 12	x_{41} 13	x_{42} 14	x_{43} 15	x_{44} 16
1	1.500															
2	.900	1.500														
3	1.200	1.200	2.000													
4	1.200	1.200	1.600	2.000												
5	.216	.216	.288	.288	1.580											
6	.216	.216	.288	.288	1.080	1.580										
7	.288	.288	.384	.384	1.440	1.440	2.520									
8	.288	.288	.384	.384	1.440	1.440	1.920	2.520								
9	.252	.252	.336	.330	.360	.360	.480	.480	1.780							
10	.252	.252	.336	.336	.360	.360	.480	.480	1.080	1.780						
11	.336	.336	.448	.448	.480	.480	.640	.640	1.440	1.440	2.720					
12	.336	.336	.448	.448	.480	.480	.640	.640	1.440	1.440	1.920	2.720				
13	.288	.288	.384	.384	.432	.432	.576	.576	.576	.576	.768	.768	2.060			
14	.288	.288	.384	.384	.432	.432	.576	.576	.576	.576	.768	.768	1.260	2.060		
15	.384	.384	.512	.512	.576	.576	.768	.768	.768	.768	1.024	1.024	1.680	1.680	2.940	
16	.384	.384	.512	.512	.576	.576	.768	.768	.768	.768	1.024	1.024	1.680	1.680	2.240	2.940
Mean	1.1	1.1	1.5	1.5	1.7	1.7	2.3	2.3	2.3	2.3	3.1	3.1	2.9	2.9	3.9	3.9
$\rho_{x_{tj}}$	.600	.600	.800	.800	.684	.684	.762	.762	.607	.607	.706	.706	.612	.612	.762	.762

$\rho_{x_{tj}}$ denotes reliability coefficient for an item j at time t.

Identification and scaling methods in second-order latent growth models

To understand how growth parameter estimates depend on identification and scaling methods, specifically, the marker-variable identification method, the effect-coding identification method, and the latent-standardization identification method, we express the growth parameters as a function of mean, variance and covariance of observed items, and item parameters.

The marker-variable identification method

The marker-variable identification method fixes the loading and intercept of one item per latent factor at a constant. This item is called the marker variable or reference variable. Figure 3a shows the marker-variable identification method for the latent factor at time 1, where the first item (x_{11}) is chosen as the marker variable. Consequently, three factor loadings and three item intercepts are freely estimated in the model under strong longitudinal MI (i.e., time-invariant factor loadings and item intercepts). A total of five parameters are estimated in the mean structure, specifically, three intercepts in τ and two means in $\boldsymbol{\kappa}' = [\kappa_L \;\; \kappa_S]$. The mean structure is identified with df = 16–5 = 11. With respect to the covariance structure, a total of 26 model parameters are freely estimated, which include

- 3 elements in the $\boldsymbol{\Phi}$ matrix for growth factors, that is, ϕ_{LL}, ϕ_{LS}, and ϕ_{SS}
- 4 elements in the $\boldsymbol{\Theta}$ matrix for factor uniqueness
- 16 elements in the $\boldsymbol{\Psi}$ matrix for item residual variance
- 3 elements in the factor loading matrix $\boldsymbol{\Lambda}$

The covariance structure is identified with df = (16×17)/2-26 = 110. The 2nd-order LGM is therefore identified containing 31 freely estimated parameters with df = 121.

With the model being identified, we now express the five growth parameters (κ_L, κ_S, ϕ_{LL}, ϕ_{LS}, and ϕ_{SS}) as a function of item information. Figure 3a shows the first item as the marker variable, but any of other three items can be the choice. Such arbitrariness in selecting marker variable is a major point we stress in the paper. Denote the marker variable and its associated parameters with a subscript r. The expectations of the marker variable at time 1 and time 2 are:

$$\begin{aligned} \mathrm{E}(x_{1r}) &= \tau_r + \lambda_r * E(\eta_1) \\ \mathrm{E}(x_{2r}) &= \tau_r + \lambda_r * E(\eta_2) \; . \end{aligned} \tag{5}$$

The expectations of the latent factor at time 1 and time 2 are:

$$\begin{aligned} E(\eta_1) &= E(\xi_L + 0 * \xi_S + \zeta_1) = E(\xi_L) = \kappa_L \\ E(\eta_2) &= E(\xi_L + 1 * \xi_S + \zeta_2) = E(\xi_L) + E(\xi_S) = \kappa_L + \kappa_S \; . \end{aligned} \tag{6}$$

Combining Equation (5) and (6) for solving κ_L and κ_S, we get

$$\boldsymbol{\kappa} = \begin{bmatrix} \kappa_L \\ \kappa_S \end{bmatrix} = \frac{1}{\lambda_r} \begin{bmatrix} E(x_{1r}) - \tau_r \\ E(x_{2r}) - E(x_{1r}) \end{bmatrix} . \tag{7}$$

Next, we derive the variance and covariance of growth factors as a function of item information. To reach this goal, equations for the marker variable and the latent factor at three time points are needed. For the first three time points:

$$\begin{aligned} x_{1r} &= \tau_r + \lambda_r \eta_1 + \varepsilon_{1r} & \quad \eta_1 &= \xi_L + 0 * \xi_S + \zeta_1 \\ x_{2r} &= \tau_r + \lambda_r \eta_2 + \varepsilon_{2r} \;\; \text{and} & \quad \eta_2 &= \xi_L + 1 * \xi_S + \zeta_2 \; . \\ x_{3r} &= \tau_r + \lambda_r \eta_3 + \varepsilon_{3r} & \quad \eta_3 &= \xi_L + 2 * \xi_S + \zeta_3 \end{aligned}$$

The covariances of the marker variable measured at the first three time points are:

$$\begin{aligned} \sigma_{x_{1r}x_{2r}} &= (\lambda_r)^2 \sigma_{\eta_1\eta_2} \\ \sigma_{x_{1r}x_{3r}} &= (\lambda_r)^2 \sigma_{\eta_1\eta_3} \; , \\ \sigma_{x_{2r}x_{3r}} &= (\lambda_r)^2 \sigma_{\eta_2\eta_3} \end{aligned} \tag{8}$$

and the covariances of the latent factor at the first three time points are:

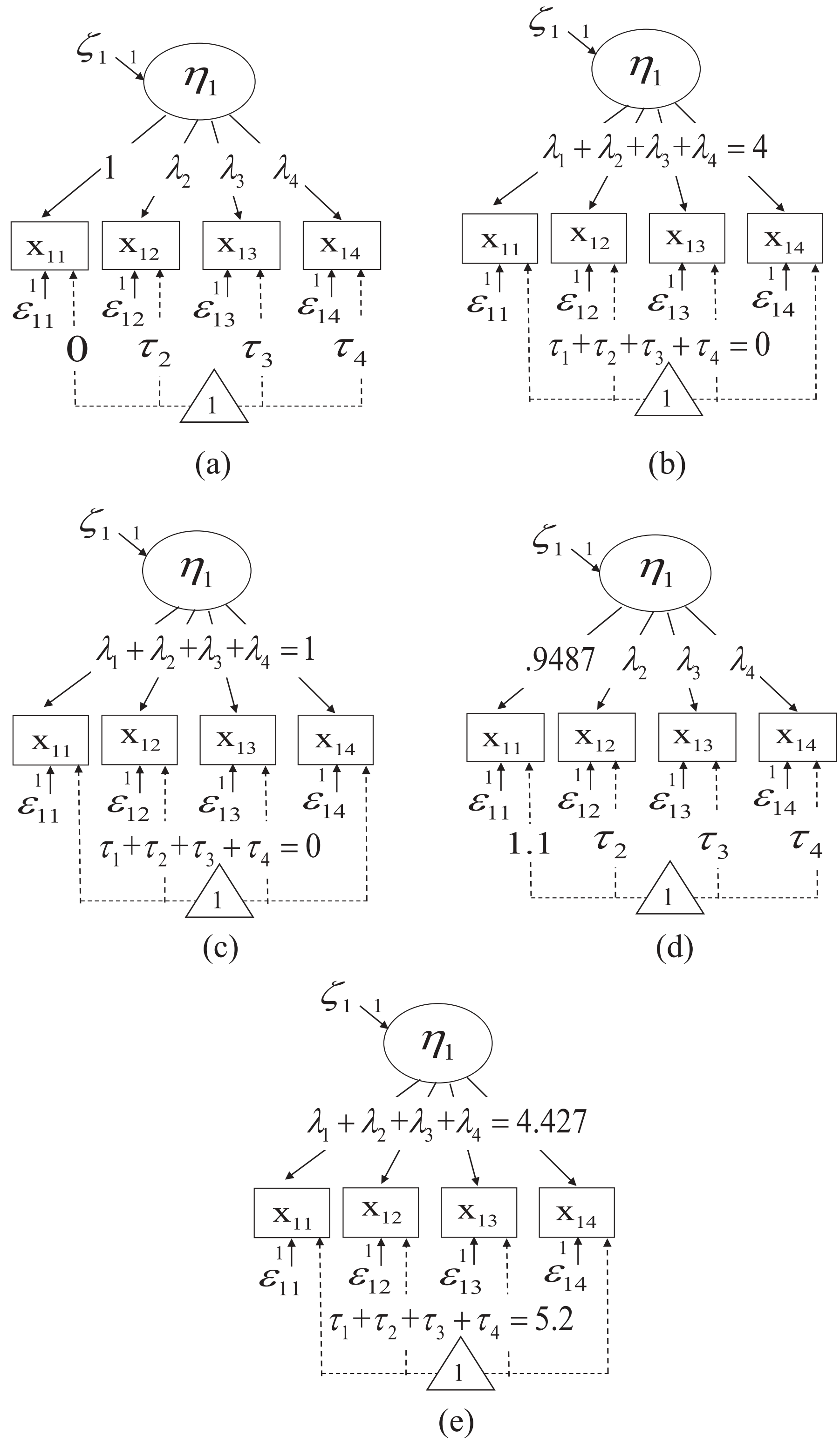

Figure 3. Marker-variable, effect-coding, and latent-standardization identification methods.
Model parameters related to the covariance structure are indicated by the solid lines; model parameters related to the mean structure are indicated by the dash lines. The other part of the 2nd-order LGM is not shown in the figure.

$$\begin{aligned}\sigma_{\eta_1\eta_2} &= \phi_{LL} + \phi_{LS} \\ \sigma_{\eta_1\eta_3} &= \phi_{LL} + 2 * \phi_{LS} \\ \sigma_{\eta_2\eta_3} &= \phi_{LL} + 3 * \phi_{LS} + 2 * \phi_{SS}\end{aligned} \quad . \tag{9}$$

Combining Equations (8) and (9) for solving ϕ_{LL}, ϕ_{LS}, and ϕ_{SS}, we get:

$$\begin{aligned}\mathbf{\Phi} &= \begin{bmatrix}\phi_{LL} & \\ \phi_{LS} & \phi_{SS}\end{bmatrix} \\ &= \frac{1}{\lambda_r^2}\begin{bmatrix}2\sigma_{x_{1r}x_{2r}} - \sigma_{x_{1r}x_{3r}} & \\ \sigma_{x_{1r}x_{3r}} - \sigma_{x_{1r}x_{2r}} & \frac{1}{2}\sigma_{x_{2r}x_{3r}} + \frac{1}{2}\sigma_{x_{1r}x_{2r}} - \sigma_{x_{1r}x_{3r}}\end{bmatrix}.\end{aligned} \tag{10}$$

Equation (7) and Equation (10) show that under strong longitudinal MI, the growth parameter estimates are not unique. They may be different if (1) different constants are assigned to λ_r and/or τ_r; or (2) a different item is chosen as the marker variable unless the loading and intercept are equal across items within a time point (i.e., longitudinal tau-equivalence assumption is satisfied; however, such an assumption is unlikely to hold in practice).

To further illustrate the dependency of growth parameter estimates on the choice of the marker variable and the constant assigned to λ_r and τ_r, we applied Equations (7) and (10) to obtain the mean, variance, and covariance of growth factors by using six marker-variable approaches listed below. The resulting growth parameter estimates are presented in Table 2. Interested readers may fit the 2nd-order LGM to the provided population covariance matrix and means to empirically verify our derivation and computation. M*plus* outputs and R scripts for all analyses discussed in the paper can be found at https://github.com/yang-fsu/second-order-LGM. The six marker-variable identification approaches are:

- Marker-1: $x_{.1}$ (i.e., the first item at each time point) was chosen as the marker variable. The loading and the intercept of this item were fixed at 1 and 0, respectively.
- Marker-2: $x_{.1}$ was chosen as the marker variable. The loading and the intercept of this item were fixed at .60 (the true value) and 0, respectively.
- Marker-3: $x_{.1}$ was chosen as the marker variable. The loading and the intercept of this item were fixed at .60 and .50 (the true values), respectively.
- Marker-4: $x_{.4}$ (i.e., the last item at each time point) was chosen as the marker variable. The loading and the intercept of this item were fixed at 1 and 0, respectively.
- Marker-5: $x_{.4}$ was chosen as the marker variable. The loading and the intercept of this item were fixed at .80 (the true value) and 0, respectively.
- Marker-6: $x_{.4}$ was chosen as the marker variable. The loading and the intercept of this item were fixed at .80 and .70 (the true values), respectively.

A different choice of marker variable and constants applied to loading and intercept would lead to different mean and covariance matrix of the latent factor over time on which the growth trajectory is examined. Based on Equations (1) and (4), the mean and variance of the latent factor at time t is:

$$E(\eta_t) = \frac{E(x_{tr}) - \tau_r}{\lambda_r}, \sigma^2_{\eta t} = \frac{\rho_{x_{tr}}\sigma^2_{x_{tr}}}{\lambda_r^2}. \tag{11}$$

When $\lambda_r = 1$ and $\tau_r = 0$, we have $E(\eta_t) = E(x_{tr})$ and $\sigma^2_{\eta_t} = \rho_{x_{tr}}\sigma^2_{x_{tr}}$. Therefore, with the Marker-1 and Marker-4 methods, the mean of latent factor at time t is equal to the mean of the marker variable at time t, and the variance of the latent factor at time t is equal to the true score variance of the marker variable at time t. Let's take the Marker-1 method as an example and refer to values in Table 1, the mean of x_{11} is 1.1, the variance of x_{11} is 1.5, and the reliability coefficient of x_{11} is .60. Therefore, $E(\eta_t)$= 1.1 and $\sigma^2_{\eta_1} = .60 \times 1.5 = .90$. This procedure can be applied when other marker-variable identification approaches are used to obtain the mean and variance of latent factor over time, which are reported in the last two columns in Table 2.

Because the latent factor is assigned a metric in different ways, the growth parameter estimates are different under different marker-variable approaches. An applied researcher who uses the best item (the item with the highest loading) as the marker variable and fixes the loading and intercept at .80 and 0 (i.e., Marker-5 approach) would conclude that "the mean of the initial status is $\hat{\kappa}_L$ =1.875 and the average linear growth rate is $\hat{\kappa}_S$ =1." Another researcher who uses the first item as the marker variable (i.e., Marker-1 approach) and fixes the loading and intercept at 1 and 0 would obtain a lower initial status (1.1) and a flatter linear growth rate (.6). Such a confusion may be avoided if the marker variable and the applied constants are explicitly stated. For example, with the Marker-1 approach, $\hat{\kappa}_L$ is the mean of the first item at time 1, and $\hat{\kappa}_S$ is the mean difference in the first item between time 2 and time 1. Assuming the growth trajectory is correctly specified, $\hat{\kappa}_S$ is also the mean difference in the first item between any two adjacent time points. As evidenced in our review, nearly all papers applied the Marker-1 approach in analyses. Returning to Knight et al. (2018) examining the development of familism values across 5th to 10th graders, the values of 4.5 and .04 are actually the estimated mean of "support" in time 1 and the estimated mean difference in "support" between adjacent time points, respectively. The values would have been different had another indicator (i.e., obligations, family as referent, and respect) been chosen as marker variable with different constants applied to the loading and intercept; the variance components and statistical significance may correspondingly be changed.

The effect-coding identification method

The effect-coding identification (ECI) method is proposed by Little et al. (2006) for which the sum of loadings is constrained to be the number of items measuring the latent factor, and the sum of item intercepts is fixed at zero. Figure 3b shows this method by using time 1 as an example. With these constraints, all items are used to establish a scale for the latent factor. From this point of view, the ECI method is referred to as a non-arbitrary method (Little et al., 2006). As we show below, however, the ECI method

Table 2. Estimated growth parameters, and mean and covariance matrix of factors using the marker-variable identification method.

	Parameter constraints	$[\hat{\kappa}_L, \hat{\kappa}_S]$	$\begin{bmatrix} \hat{\phi}_{LL} & \\ \hat{\phi}_{SS} & \hat{\phi}_{LS} \end{bmatrix}$	$\mathbf{u}_{\eta\eta}$	$\mathbf{\Sigma}_{\eta\eta}$			
Marker-1	$\lambda_{\cdot 1} = 1$	1.1	.180	1.1	.900			
	$\tau_{\cdot 1} = 0$	.6	.036 .036	1.7	.216	1.080		
				2.3	.252	.360	1.080	
				2.9	.288	.432	.576	1.260
Marker-2	$\lambda_{\cdot 1} = .6$	1.833	.5	1.833	2.500			
	$\tau_{\cdot 1} = 0$	1	.1 .1	2.833	.600	3.000		
				3.833	.700	1.000	3.000	
				4.833	.800	1.200	1.600	3.500
Marker-3	$\lambda_{\cdot 1} = .6$	1	.5	1.0	2.500			
	$\tau_{\cdot 1} = .5$	1	.1 .1	2.0	.600	3.000		
				3.0	.700	1.000	3.000	
				4.0	.800	1.200	1.600	3.500
Marker-4	$\lambda_{\cdot 4} = 1$	1.5	.320	1.5	1.600			
	$\tau_{\cdot 4} = 0$	.8	.064 .064	2.3	.384	1.920		
				3.1	.448	.640	1.920	
				3.9	.512	.768	1.024	2.240
Marker-5	$\lambda_{\cdot 4} = .8$	1.875	.5	1.875	2.500			
	$\tau_{\cdot 4} = 0$	1	.1 .1	2.875	.600	3.000		
				3.875	.700	1.000	3.000	
				4.875	.800	1.200	1.600	3.500
Marker-6	$\lambda_{\cdot 4} = .8$	1	.5	1.0	2.500			
	$\tau_{\cdot 4} = .7$	1	.1 .1	2.0	.600	3.000		
				3.0	.700	1.000	3.000	
				4.0	.800	1.200	1.600	3.500
Marker-7	$\lambda_{\cdot 1} = .9487$							
(to standardize latent factor based on x_{11})	$\tau_{\cdot 1} = 1.1$			0	1.000			
		0	.20	0.632	.240	1.200		
Marker-8	$\lambda_{\cdot 4} = 1.265$	.632	.04 .04	1.265	.280	.400	1.200	
(to standardize latent factor based on x_{14})	$\tau_{\cdot 4} = 1.5$			1.897	.320	.480	.640	1.400

$\mathbf{u}_{\eta\eta}$ and $\mathbf{\Sigma}_{\eta\eta}$ denote the mean vector and covariance matrix of latent factors over time, respectively.

is not necessarily a non-arbitrary method because loadings and item intercepts do not have to be fixed in this way. We propose two sets of constants which also render meaningful interpretations of growth parameter estimates.

When using the ECI method, the example model contains four loading parameters and four item intercepts parameters under strong longitudinal MI. With the additional constraints being placed on the sum of loadings and the sum of item intercepts, the number of freely estimated loading parameters and intercept parameters are both 3, the same as those under the marker-variable identification method. Consequently, the model is identified with df =121. To show why and how the choice of constants applied to the sum of loadings and the sum of item intercepts would lead to different growth parameter estimates, we again express growth parameters as a function of item information. We define composite variables $x_t^c = \sum_{j=1}^{J} x_{tj}$, $\tau_t^c = \sum_{j=1}^{J} \tau_{tj}$, $\lambda_t^c = \sum_{j=1}^{J} \lambda_{tj}$ and $\varepsilon_t^c = \sum_{j=1}^{J} \varepsilon_{tj}$ at time t, respectively. The superscript c denotes a quantity associated with composite scores. Under strong longitudinal MI, $\tau_t^c = \tau^c$ and $\lambda_t^c = \lambda^c$. Therefore, the equation for the composite variable x_t^c is

$$\begin{bmatrix} x_1^c \\ x_2^c \\ x_3^c \\ x_4^c \end{bmatrix} = \begin{bmatrix} \tau^c \\ \tau^c \\ \tau^c \\ \tau^c \end{bmatrix} + \begin{bmatrix} \lambda^c & 0 & 0 & 0 \\ 0 & \lambda^c & 0 & 0 \\ 0 & 0 & \lambda^c & 0 \\ 0 & 0 & 0 & \lambda^c \end{bmatrix} * \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1^c \\ \varepsilon_2^c \\ \varepsilon_3^c \\ \varepsilon_4^c \end{bmatrix} \tag{12}$$

The expectation of x_t^c at the first two time points are:

$$\begin{aligned} E(x_1^c) &= \tau^c + \lambda^c * E(\eta_1) \\ E(x_2^c) &= \tau^c + \lambda^c * E(\eta_2). \end{aligned} \tag{13}$$

The expectations of the latent factor at the first two time points have the same form as those in Equation (6). Combining Equations (13) and (6) for solving κ_L and κ_S, we get

$$\boldsymbol{\kappa} = \begin{bmatrix} \kappa_L \\ \kappa_S \end{bmatrix} = \frac{1}{\lambda^c} \begin{bmatrix} E(x_1^c) - \tau^c \\ E(x_2^c) - E(x_1^c) \end{bmatrix} \tag{14}$$

Using the information of the composite variable at the first three time points and following the same procedure as we did for the marker-variable method, the variances and covariance of the growth factors are:

$$\begin{aligned} \boldsymbol{\Phi} &= \begin{bmatrix} \phi_{LL} & \\ \phi_{LS} & \phi_{SS} \end{bmatrix} \\ &= \frac{1}{(\lambda^c)^2} \begin{bmatrix} 2\sigma_{x_1^c x_2^c} - \sigma_{x_1^c x_3^c} & \\ \sigma_{x_1^c x_3^c} - \sigma_{x_1^c x_2^c} & \frac{1}{2}\sigma_{x_2^c x_3^c} + \frac{1}{2}\sigma_{x_1^c x_2^c} - \sigma_{x_1^c x_3^c} \end{bmatrix}. \end{aligned} \tag{15}$$

Note that Equations (14) and (15) are very similar to Equations (7) and (10), except that the composite variables and their corresponding information replace those of the marker variables. Similar to the case under the marker-variable identification method, a different set of constants applied to λ^c and τ^c assigns a different metric to the latent factor. Specifically, based on Equations (12)-(13), the mean and variance of the latent factor at time t is:

$$E(\eta_t) = \frac{E(x_t^c) - \tau^c}{\lambda^c}, \sigma_{\eta_t}^2 = \frac{\sigma_{x_t^c}^2 - \psi_{x_t^c}}{(\lambda^c)^2} = \frac{\rho_{x_t^c}\sigma_{x_t^c}^2}{(\lambda^c)^2} \tag{16}$$

where $\sigma_{x_t^c}^2$, $\psi_{x_t^c}$ and $\rho_{x_t^c}$ denote composite score variance, residual variance for the composite scores, and the reliability coefficient of composite scores at time t, respectively.

Because the growth trajectory is examined on the basis of latent factors, we would expect obtaining a different set of growth parameter estimates when different values are applied to λ^c and τ^c. Therefore, the ECI method is also an arbitrary approach in terms of obtained model parameter estimates. However, three sets of selected constants for the sum of loadings and intercepts lead to meaningful interpretation. In this section, we discuss two of the sets. These two sets also provide comparable growth parameter estimates to those using 1st-order LGMs, and the scale of latent factor is directly related to the reliable component of composite scores. The third set is to standardize the latent factor at the reference time point and is therefore elaborated under the section "The Latent-Standardization Identification Method."

To illustrate, we computed composite scores as the sum of item scores and as the mean of item scores, respectively, based on the information provided in Table 1. The mean vector ($\mathbf{u}_{x^cx^c}$) and covariance matrices ($\boldsymbol{\Sigma}_{x^cx^c}$) of composite scores are provided in the last two rows in Table 3. The reliability of composite scores was .9074, .9145, .8869, and .9014 from the first to the last time points, respectively. Next, we fitted the 2nd-order LGM using the ECI method with the two sets of constants, labeled as "ECI-1" and "ECI-2" as detailed below.

- ECI-1: The sum of loadings was fixed at 4 and the sum of the item intercepts was fixed at 0 (Figure 3b).
- ECI-2: The sum of loadings was fixed at 1 and the sum of the item intercepts was fixed at 0 (Figure 3c).

These two approaches assigned different metrics to the latent factor. To demonstrate it, we applied these two sets of constants to Equation (16). The ECI-1 approach resulted in $E(\eta_1) = (1.1 + 1.1 + 1.5 + 1.5)/4 = 1.3$ and $\sigma^2_{\eta 1} = .9074 \times 21.60/16 = 1.225$, and the ECI-2 approach yielded $E(\eta_1) = 1.1 + 1.1 + 1.5 = 5.2$ and $\sigma^2_{\eta 1} = .9074 \times 21.60 = 19.60$. In words, the mean of the latent factor at time 1 is equaled to the mean of *item mean scores* (the ECI-1 approach) or the mean of *summed item scores* (the ECI-2 approach) at time 1. The variance of the latent factor at time 1 is equaled to the true score variance of the *item mean scores* (the ECI-1 approach) or the *summed item scores* (the ECI-2 approach) at time 1. The same procedure can be applied to other time points to obtain mean and variance of latent factor at time 2 to time 4.

A different scaling of latent factor leads to a different set of growth parameter estimates. Based on Equation (14), the ECI-1 approach resulted in $\hat{\boldsymbol{\kappa}} = [1.3, .7]$ and the ECI-2 approach resulted in $\hat{\boldsymbol{\kappa}} = [5.2, 2.8]$. In words, under the ECI method, the composite variable at time 1 is represented by the *item mean scores* (the ECI-1 approach) or the *summed item scores* (the ECI-2 approach) at time 1. Consequently, $\hat{\kappa}_L$ is the estimated mean of *item mean scores* (the ECI-1 approach) or *summed item scores* (the ECI-2 approach) at time 1, and $\hat{\kappa}_S$ estimated is the mean difference in *item mean scores* (the ECI-1 approach) or *summed item scores* (the ECI-2 approach) between time 2 and time 1. Assuming the growth trajectory is correctly specified, $\hat{\kappa}_S$ is also the mean difference in *item mean scores* (the ECI-1 approach) or *summed item scores* (the ECI-2 approach) between any two adjacent time points. The variance components ($\hat{\phi}_{LL}$, $\hat{\phi}_{SS}$, and $\hat{\phi}_{LS}$) and the statistical testing would be based on covariances of the corresponding composite scores over time as shown in Equation (15).

Connecting to a common practice of using composite scores to apply a 1st-order LGM, the ECI-1 approach would lead to comparable results to a 1st-order LGM where *item mean scores* are used to represent the construct (labeled as "1st-order with means"); the ECI-2 approach would lead to comparable results to a 1st-order LGM where *summed item scores* are used to represent the construct (labeled as "1st-order

Table 3. Estimated growth parameters, and mean and covariance matrix of factors using the effect-coding identification method and 1st-order LGMs.

	Parameter constraints	$[\hat{\kappa}_L, \hat{\kappa}_S]$	$\begin{bmatrix}\hat{\phi}_{LL} & \\ \hat{\phi}_{LS} & \hat{\phi}_{SS}\end{bmatrix}$	$\mathbf{u}_{\eta\eta}$	$\boldsymbol{\Sigma}_{\eta\eta}$			
ECI-1 (sum of 4)	$\sum_{j=1}^{4}\lambda_{tj} = 4$ $\sum_{j=1}^{4}\tau_{tj} = 0$	$\begin{bmatrix}1.3\\ .7\end{bmatrix}$	$\begin{bmatrix}.245 & \\ .049 & .049\end{bmatrix}$	1.3	1.225			
				2.0	.294	1.470		
				2.7	.343	.490	1.470	
				3.4	.392	.588	.784	1.715
ECI-2 (sum of 1)	$\sum_{j=1}^{4}\lambda_{tj} = 1$ $\sum_{j=1}^{4}\tau_{tj} = 0$	$\begin{bmatrix}5.2\\ 2.8\end{bmatrix}$	$\begin{bmatrix}3.92 & \\ .784 & .784\end{bmatrix}$	5.2	19.600			
				8.0	4.704	23.520		
				10.8	5.488	7.840	23.520	
				13.6	6.272	9.408	12.544	27.440
ECI-3 (to standardize latent factor)	$\sum_{j=1}^{4}\lambda_{tj} = 4.427$ $\sum_{j=1}^{4}\tau_{tj} = 5.2$	$\begin{bmatrix}0\\ .632\end{bmatrix}$	$\begin{bmatrix}.20 & \\ .04 & .04\end{bmatrix}$	0	1.000			
				.632	.240	1.200		
				1.265	.280	.400	1.200	
				1.897	.320	.480	.640	1.400
				$\mathbf{u}_{x^cx^c}$	$\boldsymbol{\Sigma}_{x^cx^c}$			
1st-order with means		$\begin{bmatrix}1.3\\ .7\end{bmatrix}$	$\begin{bmatrix}.245 & \\ .049 & .049\end{bmatrix}$	1.3	1.350			
				2.0	.294	1.6075		
				2.7	.343	.490	1.6575	
				3.4	.392	.588	.784	1.9025
1st-order with sums		$\begin{bmatrix}5.2\\ 2.8\end{bmatrix}$	$\begin{bmatrix}3.92 & \\ .784 & .784\end{bmatrix}$	5.2	21.60			
				8.0	4.704	25.72		
				10.8	5.488	7.84	26.52	
				13.6	6.272	9.408	12.544	30.44

$\mathbf{u}_{\eta\eta}$ and $\boldsymbol{\Sigma}_{\eta\eta}$ denote the mean vector and covariance matrix of latent factors over time, respectively. $\mathbf{u}_{x^cx^c}$ and $\boldsymbol{\Sigma}_{x^cx^c}$ denote the mean vector and covariance matrix of composite scores over time, respectively.

with sums"). To illustrate similarities and differences between the 1st-order LGM and the 2nd-order LGM with ECI methods, we conducted these two 1st-order LGMs using our illustrative example. As expected, the two compared models yielded identical estimates of five growth parameters (see also Leite, 2007 for a comparison of 1st-order LGM and 1st-order LGM with fixed error variance for composite scores; the latter case gives the same results as the ECI method). This is not surprising because the means and covariances of factor scores obtained from the ECI method were equaled to the means and covariances of composite scores obtained from the 1st-order LGMs (e.g., both "ECI-1" and "1st-order with means" yielded the mean vector of [1.3, 2.0, 2.7, 3.4]). Note that the variance of composite scores (i.e., the diagonal elements of covariance matrix) was greater than the corresponding variance of factor scores from the 2nd-order LGM. This is because the 1st-order LGM failed to take measurement errors into account. With sample data, we expect that applying the 1st-order LGM would lead to overestimation of standard errors associated with growth parameter estimates and consequently lower power for statistical testing of the growth parameters.

The latent-standardization identification method

The latent-standardization identification method is to place the latent factor at the reference time point (time 1 in our example; η_1) on a standardized metric such that $\eta_1 \sim N(0,1)$ (Ferrer et al., 2008; Grimm et al., 2017). Consequently, the mean of the slope factor (κ_S) is the mean change scores between two adjacent time points relative to the standard deviation of scores at the reference time point. Glass et al. (1981) labeled this standardized mean difference as *Glass's delta*.[5] Different from multiple-group factor analysis, standardization of η_1 in a 2nd-order LGM is not straightforward because mean and variance of η_1 are not model parameters. Standardization of η_1 is achieved by placing constraints on other model parameters. Below we show three approaches to achieving latent standardization through the marker-variable and ECI methods. In other words, the latent-standardization identification method can be viewed as a special case of the marker-variable identification and ECI methods.

To achieve latent standardization through the marker-variable identification method, we revisit Equation (11) and place two constraints on the intercept and loading of the marker variable: (1) fixing the intercept of the marker variable at time 1 as the mean of the marker variable, that is, $\tau_r = \mathrm{E}(\mathrm{x}_{1r})$, and (2) fixing the loading of the marker variable at time 1 as the square root of the true score variance of the marker variable, that is, $\lambda_r = \sqrt{\rho_{\mathrm{x}_{1r}}\sigma^2_{\mathrm{x}_{1r}}}$. This approach is shown in Figure 3d where the first item was chosen as the marker variable. Consequently, the latent factor η_1 has a mean of zero and standard deviation of one.

To achieve latent standardization through the ECI method, we revisit Equation (16) and place two constraints on the sum of intercepts and the sum of loadings at time 1: (1) fixing the sum of intercepts as the sum of item means at time 1, that is, $\tau^c = \mathrm{E}(\mathrm{x}^c_1)$, and (2) fixing the sum of loadings as the square root of the true composite score variance at time 1, that is, $\lambda^c = \sqrt{\rho_{\mathrm{x}^c_1}\sigma^2_{\mathrm{x}^c_1}}$. This approach is shown in Figure 3e. Consequently, the latent factor η_1 would have a mean of zero and standard deviation of one.

To illustrate the use of the marker-variable and ECI methods to standardize latent factor at time 1, we conducted the following three analyses using the numerical example.

- Marker-7 (standardization based on the first item): $\mathrm{x}_{\cdot 1}$ was chosen as the marker variable. Its loading and intercept were fixed at $\sqrt{\rho_{\mathrm{x}_{11}}\sigma^2_{\mathrm{x}_{11}}} = \sqrt{.6 \times 1.5} = .9487$ and 1.1, respectively.
- Marker-8 (standardization based on the last item): $\mathrm{x}_{\cdot 4}$ was chosen as the marker variable. Its loading and intercept were fixed at $\sqrt{\rho_{\mathrm{x}_{14}}\sigma^2_{\mathrm{x}_{14}}} = \sqrt{.8 \times 2.0} = 1.265$ and 1.5, respectively.
- ECI-3: the sum of loadings and the sum of intercepts were fixed at $\sqrt{\rho_{\mathrm{x}^c_1}\sigma^2_{\mathrm{x}^c_1}} = \sqrt{.9074 \times 21.6} = 4.427$ and $E(\mathrm{x}_{11} + \mathrm{x}_{12} + \mathrm{x}_{13} + \mathrm{x}_{14}) = 5.2$, respectively.

The results were shown in the last two rows in Table 2 and the last row in Table 3. As expected, these three analyses yielded the same growth parameter estimates because the latent factor η_1 is standardized. The mean of the slope factor ($\hat{\kappa}_S$) was .632 and was interpreted as .632 standard deviation increase in scores from one time point to the next time point. Grimm et al. (2017) illustrated a slightly different approach to standardize the latent factor at the reference time point with an empirical dataset. Instead of placing a constraint on the sum of item intercepts, Grimm et al. (2017) fixed the mean of level factor (κ_L) at zero, while the sum of loadings was fixed in the same way as we had in the ECI-3 approach (see pp. 361–367).

Through demonstration, we can conclude that growth parameter estimates are unique under the latent-standardization identification method, although the standardization of the latent factor at the reference time point can be achieved in multiple ways. For this reason, the latent-standardization identification method may be preferred by researchers who are interested in interpreting the mean of the slope factor (the mean of the level factor is always 0).

Conclusion and discussion

Using both formulas and a numerical example, we show that growth parameter estimates in 2nd-order LGMs are not unique; they depend on identification and scaling methods applied to latent factors, upon which the growth trajectory is tested. Although we discuss identification and scaling issues using an example of linear latent growth models, our

[5]Glass's delta defined by Glass et al. (1981, p. 29) is equaled to the mean difference between the treatment and control groups divided by the standard deviation in the control group. Glass's delta is different from Cohen's *d* effect size. In Cohen's *d*, the denominator is the pooled standard deviation from the treatment and control groups.

derivation procedure can be extended to 2^{nd}-order LGMs with different growth trajectories (e.g., a quadratic shape).

We show only 11 ways of identification (Marker-1 to Marker-8 and ECI-1 to ECI-3) using a numerical example, but our derivation indicates that the marker-variable and ECI methods can be applied in infinite ways because any item can be chosen as the marker variable and there are infinite ways to apply the constants to its loading and intercept (or the sum of loadings and the sum of intercepts). A different choice of marker variable and constants may lead to very different interpretations of growth trajectory. On the other hand, although latent standardization can be achieved in multiple ways, the latent-standardization identification method leads to a unique set of growth parameter estimates and may be preferred by researchers. Regardless of which identification and scaling method is chosen, we suggest users of 2^{nd}-order LGMs state explicitly how the latent factors are scaled, and incorporate this information when interpreting growth parameter estimates.

Perhaps because of the default setting in software programs (e.g., M*plus*), the marker-variable identification method with the first item being the marker variable is often applied in empirical studies. Such a default setting is not an issue when the first item has sound psychometric properties (e.g., standardized loading is reasonably large and satisfies longitudinal MI). A better choice is to select the best item (e.g., the item 3 and item 4 in our numerical example) as the marker variable. The ECI methods do not scale the latent factor using only one item; instead, it scales the latent factor in the unit of reliable component of all items. For the ECI methods, we highlight three sets of constraints that lead to meaningful interpretation of growth trajectories. One set (sum of item loadings =number of the items; sum of item intercepts =0) is proposed by Little et al. (2006), which is analogous to the use of first-order latent growth model with *mean item scores* being the approximation of latent construct. The second set (sum of item loadings =1; sum of item intercepts =0) is analogous to the use of 1^{st}-order LGM with *summed item scores* being the approximation of latent factors. The third set leads to standardization of the latent factor at the reference time point. Applied researchers are encouraged to make their own choice when analyzing longitudinal data. As has been stressed in the paper, it is important to explicitly state the identification and scaling method to avoid misinterpretation of model results when a 2^{nd}-order LGM is applied.

Our derivation and illustration are based on the assumption of strong factorial invariance. It is well discussed in the multiple-group structural equation modeling and longitudinal modeling that when the strong factorial invariance assumption is not met, latent factors are not the cause of changes in observed item scores over time or across groups (e.g., Little et al., 2007; Meredith, 1993; Stoel et al., 2004; Thompson & Green, 2013). Consequently, the resulting growth parameter estimates may not be meaningful. In practice, it is not uncommon that this assumption is only partially met. In these circumstances, researchers may adopt partial MI models. However, different problems may arise when fitting models with partial MI under the marker-variable identification method and the ECI method. For the marker-variable identification method, Ferrer et al. (2008) and Stoel et al. (2004) demonstrated that under partial MI using different marker variables did not lead to the same expected growth trajectory (e.g., a linear growth trend). For the ECI method, growth parameter estimates are expected to be accurate when the longitudinal MI is met at the composite variable level (see Equation 12). However, when the sum of loadings or the sum of intercepts varies over time, the resulting growth trajectory may not be meaningful. Future studies can be conducted to evaluate the relative performance of the marker-variable identification method and the ECI method under various partial measurement invariance conditions in 2^{nd}-order LGMs.

ORCID

Yanyun Yang http://orcid.org/0000-0001-6671-1583

References

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Wiley.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421–483. https://doi.org/10.1177/109442819814004

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*, 22–36. https://doi.org/10.1027/1614-2241.4.1.22

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Sage.

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Press.

Hancock, G. R., Kuo, W.-L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, *8*, 470–489. https://doi.org/10.1207/S15328007SEM0803_7

Klößner, S., & Klopp, E. (2019). Explaining constraint interaction: How to interpret estimated model parameters under alternative scaling methods. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 143–155. https://doi.org/10.1080/10705511.2018.1517356

Knight, G. P., Mazza, G. L., & Carlo, G. (2018). Trajectories of familism values and the prosocial tendencies of Mexican American adolescents. *Developmental Psychology*, *54*, 378–384. https://doi.org/10.1037/dev0000436

Laenen, A., Alonso, A., Molenberghs, G., & Vangeneugden, T. (2009). Reliability of a longitudinal sequence of scale ratings. *Psychometrika*, *74*, 49–64. https://doi.org/10.1007/s11336-008-9079-7

Leite, W. L. (2007). A comparison of latent growth models for constructs measured by multiple items. *Structural Equation Modeling*, *14*, 581–610. https://doi.org/10.1080/10705510701575438

Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 121–147). LEA.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, *13*, 59–72. https://doi.org/10.1207/s15328007sem1301_3

Marcoulides, K. M. (2019). Reliability estimation in longitudinal studies using latent growth curve modeling. *Measurement: Interdisciplinary Research and Perspectives*, *17*, 67–77. https://doi.org/10.1080/15366367.2018.1522169

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

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

Robinson, K. A., Perez, T., Nuttall, A. K., Roseth, C. J., & Linnenbrink-Garcia, L. (2018). From science student to scientist: Predictors and outcomes of heterogeneous science identity trajectories in college. *Developmental Psychology*, *54*, 1977–1992. https://doi.org/10.1037/dev0000567

Stoel, R. D., van den Wittenboer, G., & Hox, J. (2004). Methodological issues in the application of the latent growth curve model. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models* (pp. 241–261). Kluwer Academic Publishers.

Thompson, M. S., & Green, S. B. (2013). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling* (2nd ed., pp. 163–218). Information Age Publishing.

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development*, *89*, e138–e156. https://doi.org/10.1111/cdev.12806

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, *49*, 587–602. https://doi.org/10.1177/001316448904900309