The Dissertation Committee for Walter Lana Leite

certifies that this is the approved version of the following dissertation:

A Comparison of Latent Growth Models for Constructs Measured by Multiple Indicators

Committee:

Laura M. Stapleton, Supervisor

S. Natasha Beretvas

Keenan A. Pituch

William R. Koch

Thomas W. Pullum

A Comparison of Latent Growth Models for Constructs Measured by Multiple Indicators

by

Walter Lana Leite, B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the degree of

Doctor of Philosophy

The University of Texas a Austin

May 2005

Dedicated to my parents

Nelson José Pereira Leite and

Maria do Rosário de Lana Leite,

my sisters Janise and Flávia,

and my first academic mentor, Iris Goulart

Acknowledgements

First, I'd like to thank the two professors that have helped with my journey through graduate school from the moment I joined The University of Texas at Austin until the moment I completed my dissertation: Laura Stapleton and Tasha Beretvas. Laura, as my supervisor, has always been available to talk to me about my dissertation and her excellent feedback allowed my initial research ideas to grow into a complete dissertation. I am also very thankful for learning structural equation modeling with her, which not only has provided a very interesting topic for my dissertation but has also indicated the initial direction of my academic career. Tasha is the person without whom I would not have become a quantitative methods student. I need to thank her for offering me the opportunity to do research with her when I was an exchange student at UT, and also for giving me the incentive to join the quantitative methods program as a doctoral student. I am in debt with Tasha for working with me in my first publication, which was a very important moment in my career. I am also grateful for the feedback Tasha has provided me during my dissertation. Laura and Tasha have played the greatest part in educating me into the quantitative methodologist I have become, and I am deeply thankful for having the privilege to know and work with them.

I'd like to thank the members of my committee (Laura Stapleton, Tasha Beretvas, Keenan Pituck, Bill Koch, and Tom Pullum) for their help during my dissertation. Thanks to them, my dissertation writing process went as smooth as it did.

I am very grateful for all the insightful talks I've had with Barbara Dodd, which gave me a very good idea of what an academic career is all about. She is a very caring person who has always being very attentive to me. I am also thankful to her, Laura, and Tasha for their assistance in my successful academic job search.

I'd like to thank my friends Belinda, Lisa, Karen, Marcos, and Rick for their important role of providing me with some relief from the work on my dissertation. Finally, I'd like to thank my sister Flávia for being very patient with me while I was focused on writing my dissertation.

A Comparison of Latent Growth Models for Constructs Measured by Multiple Indicators

Publication No. ________________

Walter Lana Leite, Ph.D.
The University of Texas at Austin, 2005

Supervisor: Laura M. Stapleton

Latent growth modeling (LGM) of composites of multiple items (for example, means or sums of items) has been frequently used to analyze the growth of latent constructs. However, composites are only equivalent to latent constructs if the items' factor loadings are equal to one and there is no measurement error (Bollen & Lennox, 1991). In this study, the adequacy of using univariate LGM to model composites of multiple items, as well three other alternative methods were evaluated through a Monte Carlo simulation study. The four methods evaluated in this study were the univariate LGM, the univariate LGM with fixed error variances, the univariate LGM with the correction for attenuation, and the curve-of-factors model (McArdle, 1988; Tisak and Meredith, 1990).

This simulation study manipulated the number of items per construct, the number of measurement times, the sample size, the reliability of the composites, the invariance of item parameters, and whether the items were essentially tau-equivalent or essentially congeneric. One thousand datasets were simulated for each of the conditions.

The results indicate that using univariate LGM with composites of multiple items only produces unbiased parameter estimates and standard errors if the items are essentially tau-equivalent. The univariate LGM with fixed error variances performed identically to the univariate LGM. The univariate LGM with the correction for attenuation produced unbiased parameter estimates when the items were essentially tau-equivalent, but produced negatively biased estimates of standard errors.

The curve-of-factors model was found to be the most appropriate method to analyze the growth of latent constructs measured by multiple items. The curve-of-factors model was able to provide unbiased parameter estimates and standard errors under all conditions evaluated in this study. However, with sample sizes of 100 or 200, a large percentage of chi-square statistics were positively biased and the fit indices indicated inadequate model fit.

This study’s recommendation is that the curve-of-factors model should be preferred to analyze the growth of latent variables measured by multiple items, but the use of sample sizes larger than 200 is strongly recommended to help ensure that adequate fit statistics and fit indices are obtained for appropriate models.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

The study of human change across time can be a major concern not only of researchers, but of society as a whole. For example, educational accountability systems of school districts and states are required by federal law to show that there is change in the student's educational performance from grade to grade (U.S. Department of Education, 2003). Because the assessment of change is very important in our society, longitudinal data have been frequently collected by research institutions such as the National Center for Education Statistics and the National Science Foundation. For example, the data of the National Education Longitudinal Study (NELS) collected by the National Center for Education Statistics have been used in longitudinal studies about the effects of parental involvement (Fan, 2001), teacher education (Hill, 2003), and high school employment (Quirk, Keith, & Quirk, 2001) on student educational achievement.

There are several methods to study longitudinal data, such as repeated measures analysis of variance (Maxwell & Delaney, 1990), hierarchical linear models (Raudenbush & Bryk, 2002), and autoregressive models (Curran & Bollen, 2001; Hertzog & Nesselroade, 2003; Lawrence & Hancock, 1998). These techniques provide useful information about growth across time, but each of them have limitations (Hancock, Kuo, & Lawrence, 2001) that make them appropriate for certain types of research questions but not for others. For example, autoregressive models are indifferent to the functional form of change over time (Hertzog & Nesselroade, 2003), and repeated measures ANOVA cannot handle heterocedasticity and dependence between measurement errors (Raykov &

Marcoulides, 2000; Willett & Sayer, 1994). Latent growth modeling (LGM) has recently emerged as a very general method of investigating longitudinal data. The generality of this method comes from the fact that it can analyze longitudinal changes in means, variances, and covariances at both individual and group levels, and it can handle complex covariance structures, heterocedasticity and correlated errors (Holt & Stuckey, 2004; McArdle & Epstein, 1987; Raykov & Marcoulides, 2000). Because of its ability to fit complex models, LGM has been increasingly used by researchers. A search in Academic Search Premier, ERIC, and PsycInfo for the words “latent growth” in the literature published from January 2000 to December 2004 resulted in 187 articles, demonstrating that LGM has become a popular method of longitudinal data analysis. From these articles, 38 were methodological, and the other 149 were applications.

Large longitudinal surveys usually contain data on multiple indicators that can be used to model the development of latent constructs (Muthén, B. O. & Khoo, 1998). For example, the NELS contains several variables that can be used as indicators of student motivation (Singh, Granville, & Dika, 2002), parental involvement (Fan, 2001) and self-concept (Brown, 1999). MacCallum, Kim, Malarkey, and Kiecolt-Glaser (1997) pointed out that conventional methods of longitudinal analysis and the vast majority of published longitudinal analyses have focused on a single outcome variable, which has also been the case in the applied literature employing LGM. Several of the applied studies identified in the literature search mentioned above had collected longitudinal data about a latent construct using multiple indicators. However, many of these studies fitted a univariate LGM to their data, despite the latent nature of the variable of interest. Using univariate

LGM implies that the outcome variable is an observed variable, which may be adequate for variables such as cigarette use (Hser, 2001), that may have small and probably negligible measurement error. However, several of these studies assessed change in a latent construct, as measured by a multiple-item scale, and used univariate LGM to analyze composite scores based on the scales' items (Baer & Schmitz, 2000; Crawford, Cohen, Johnson, Sneed, & Brook, 2004; DuBois et al., 2002). The most common composite scores found in the literature were the means of the scales' items (Buist, Dekovi Cacute, Meeus, & van Aken, 2002; Johnson, 2002; Li, F. et al., 2001; Mason, 2001; Roth, Haley, Owen, Clay, & Goode, 2001; Willett & Keiley, 2000). The sum of the item scores was also used (Chan, Ramey, Ramey, & Schmitt, 2000; Meeus, Branje, & Overbeek, 2004). The problem of this common practice of using univariate LGM with composites of multiple indicators is that it ignores the measurement errors of the indicators. Although most studies reviewed reported reliability estimates for the scales being used, alternative models that control for measurement error such as adjusting the variance/covariance matrix for unreliability (Fan, 2003b; Lomax, 1986; Muchinsky, 1996) or fixing the composite's error variance using its reliability estimate (Bollen, 1989) were not employed.

The issue of measurement error in longitudinal analysis of multiple-item scales can also be addressed by using a multivariate extension of LGM known as the curve-of-factors model (McArdle, 1988), or second-order LGM (Hancock et al., 2001; Sayer & Cumsille, 2001). This model controls for measurement error by simultaneously estimating the measurement model across different measurement times and identifying longitudinal changes in the latent variable.

An additional limitation of univariate LGM that has been ignored in the applied literature is that univariate LGM of multiple-item composites assumes strict factorial invariance, which means that the factor loadings, error variances, and intercepts of the items are equivalent at different times of measurement (Hertzog & Nesselroade, 2003; Meredith, 1993). This is a very restrictive assumption that may not be true in the data being analyzed. When this assumption is not met, bias in the parameter estimates may arise because univariate LGM of multiple-item composites does not distinguish between changes in the scale measuring the construct and true longitudinal changes in the construct (Sayer & Cumsille, 2001).

The curve-of-factors model has the advantage over univariate LGM of being able to freely estimate the indicators' factor loadings (with the exception of one loading which is fixed to identify the model), error variances and intercepts, which allows that changes in the latent construct within measurement times be examined separately from changes across measurement times. Because the curve-of-factors model allows the estimation of the item-level parameters, statistical tests of factorial invariance can be performed. Despite the potential advantages of the curve-of-factors model for LGM of multiple-indicator composites, the literature search conducted revealed that the curve-of-factors model has only been used in a couple of applied studies (Duncan, S. C. & Duncan, 1996; Hancock et al., 2001). Furthermore, guidelines with respect to the situations in which it should be preferred over univariate LGM and the samples sizes needed for parameter estimation have not been established yet.

This study compares the performance of different methods of fitting latent growth models in the analysis of multiple-item longitudinal data. More specifically, it compares three methods of univariate LGM (composite scores, composite scores corrected for unreliability, composite scores with fixed error variances) and the curve-of-factors model. The major interest is determining in which conditions each of these methods performs adequately in the presence of measurement error. The effect of measurement non-invariance over time on the estimates from each of these methods will also be investigated, as well as sample sizes necessary for estimation. It is hoped that the results of this study will inform applied researchers who are considering the use of LGM about the adequacy of each model for their longitudinal data.

Chapter II: Literature Review

This study investigates two problems that can interfere with the quality of inferences made through the use of latent growth models with data from instruments containing multiple indicators: measurement error and construct non-invariance over time. This chapter first introduces the latent growth model and describes its characteristics and limitations. Second, the problem of measurement error and the solutions that will be attempted in this study are described. Third, the assumption of measurement invariance in longitudinal research is explained, as well as the difficulties that arise when testing it statistically. Finally, because parameter estimation and assessment of fit in latent growth models depend on sample size, the issue of determining a sufficient sample size is reviewed.

Latent growth models

Latent growth models analyze longitudinal changes in means, variances, and covariances of variables. These statistics are summarized and used in the analysis in the form of a covariance-mean (Raykov & Marcoulides, 2000) or moment matrix (McArdle, 1988). This matrix is created by attaching a row and a column vector of means to the variance/covariance matrix between *k* variables (McArdle, 1988). A constant equal to one is added at the intersection of this row and column, which increases the order of the *k x k* covariance matrix to (*k* + 1 x *k* + 1). The matrix *M* presented below is an example of moment matrix between four observed variables:

$$M = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & s_{14} & \bar{x}_1 \\ s_{12} & s_2^2 & s_{23} & s_{24} & \bar{x}_2 \\ s_{13} & s_{23} & s_3^2 & s_{34} & \bar{x}_3 \\ s_{14} & s_{24} & s_{34} & s_4^2 & \bar{x}_4 \\ \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \bar{x}_4 & 1 \end{bmatrix} \quad (1)$$

In latent growth modeling, the longitudinal changes for a group of individuals are explained by three latent variables: level, shape and error (McArdle & Hamagami, 1991). The level factor presents the status of the individuals in the observed variables at any point of the development chosen as a reference. The point of reference is defined through the coding of the loadings of the shape factor (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). Further details about ways of coding the loadings will be given later in this section. If the first measurement time is used as the reference point of development, the level can be interpreted as an intercept (Muthén, B. O. & Khoo, 1998). In this case, the intercept is interpreted as attained status on the trait due to previous experiences. If the researcher prefers to specify as the reference point a measurement occasion other than the first (e.g. the last measurement time), the level is interpreted as the status of the individuals at that specific measurement time. Each individual has a level score that is constant across all measurement times (McArdle & Hamagami, 1991).

The shape factor indicates individual differences in the trajectory of growth (Rovine & Molenaar, 1998). If the growth is linear, the shape variable can be more easily interpreted as the slope, which is the rate of change of individuals across time (Duncan, T. E., Duncan, Strycker, Li, & Alpert, 1999; McArdle & Hamagami, 1991; Muthén, B.

O. & Khoo, 1998). Each individual's shape score represents the amount of the individual is expected to change between measurement times (McArdle & Hamagami, 1991).

The error variable represents variability not related to the latent trait being measured. It is a combination of random measurement error and systematic unique factors. The measurement error scores are assumed to have a mean of zero and zero correlations with the other variables over time (McArdle & Hamagami, 1991), while the unique factor scores may be correlated over time. The error variables at each measurement time may be allowed to correlate with each other, which accounts for change in the unique factors across time. A latent growth model with three measurement times is presented in figure 1. The dotted lines indicate that the model can be expanded to include more measurement times and correlated error variables.

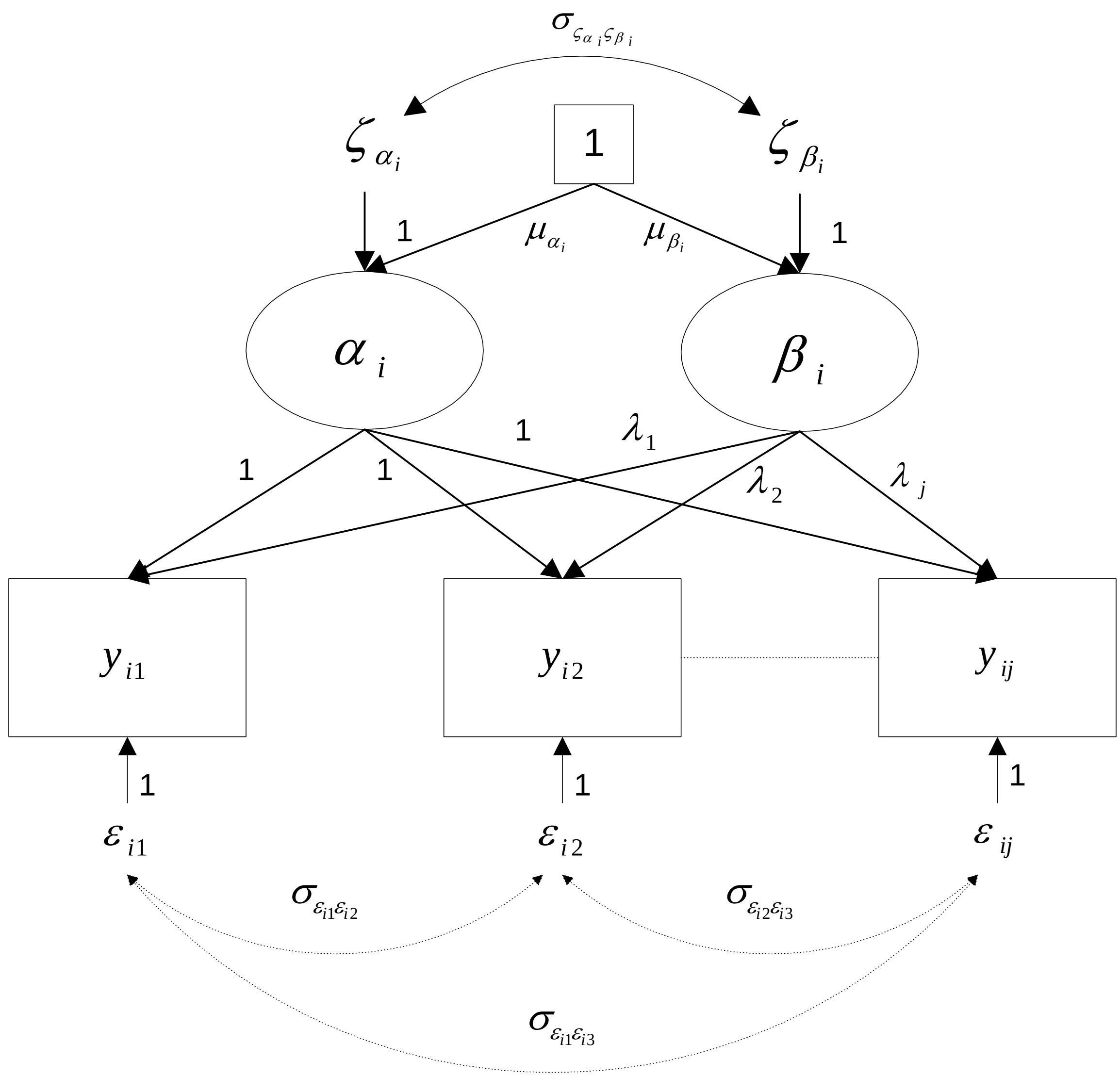

Figure 1. Latent growth model with no predictors

The univariate LGM is similar to a common factor model. The measurement part of the model can be expressed as (Singer & Willett, 2003):

$$y = \Lambda_y \eta + \varepsilon \qquad (2)$$

where y is a vector of observed scores for individual i on measurement times one to j, Λ_y is a j x 2 matrix containing a column of level coefficients and a column of shape coefficients for j measurement times, η is a 2 x 1 vector containing latent scores for the level factor, α_i, and shape factor, β_i, and ε is a vector of errors in each of the j measurement occasions.

In matrix format, this model can be expressed as:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ \cdot \\ \cdot \\ \cdot \\ y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \\ 1 & \lambda_3 \\ 1 & \lambda_4 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \lambda_j \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{ij} \end{bmatrix} \quad (3)$$

In this model, Λ_y is equivalent to a matrix of factor loadings in a common factor model. The column of ones corresponds to the loadings of the level factor. The column $\lambda_1 ... \lambda_j$ contains the factor loadings of the shape factor. In order to estimate this two factor model, two of the factor loadings of the shape factor need to be fixed (McArdle & Hamagami, 1991). Fixing one of the shape's factor loadings at zero defines its correspondent occasion of measurement as a reference (Rovine & Molenaar, 1998). As a consequence, latent means of the level and shape factors will be the expected values of these factors at the measurement time whose loading was fixed to zero. It is necessary to fix another factor loading to a non-zero value to set the scale of the shape factor. It is possible to freely estimate the other factor loadings of the shape factor, as long as the model remains identified, or fix them to any value. The differences between the shape's factor loadings can be interpreted as the number of units of change that occur between two occasions of measurement (McArdle & Hamagami, 1991).

The error vector ε is distributed with zero mean vector and covariance matrix Θ_ε. Some methods for longitudinal analysis, such as repeated measures analysis of variance, require the assumption of homocedasticity and independence of errors. In the context of univariate LGM, homocedasticity means that error variances are equal across testing times. Independence of errors means that correlations between errors across testing times is zero. In LGM, the assumption of homocedasticity can be relaxed by estimating freely the error variances. For example, the variance/covariance matrix of errors Θ_ε for a model with six measurement times allowing heterocedasticity is:

$$\Theta_\varepsilon = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6^2 \end{bmatrix} \quad (4)$$

where σ_j^2 is the error variance of the observed variable at measurement time *j*.

In LGM, the researcher has the option of assuming homocedasticity by constraining the variances to be identical. If homocedasticity is assumed, the variances along the diagonal of the error matrix above will be equivalent.

Furthermore, the assumption of independence of residuals can be relaxed by allowing correlated errors. Modeling correlated errors implies a hypothesis that there is part of the error variability that is non-random, and that is not accounted for by the latent factors. For example, the variance/covariance matrix of error for a model with six measurement times allowing errors adjacent in time to correlate is:

$$\Theta_{\varepsilon} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & 0 & 0 & 0 & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & 0 & 0 & 0 \\ 0 & \sigma_{23} & \sigma_3^2 & \sigma_{34} & 0 & 0 \\ 0 & 0 & \sigma_{34} & \sigma_4^2 & \sigma_{45} & 0 \\ 0 & 0 & 0 & \sigma_{45} & \sigma_5^2 & \sigma_{56} \\ 0 & 0 & 0 & 0 & \sigma_{56} & \sigma_6^2 \end{bmatrix} \qquad (5)$$

Specific hypotheses about the shape of the growth can be tested by fitting models where all the shape's factor loadings are fixed. For example, the researcher can choose to fix all the shape's loadings to be equal to the difference between the real measurement times, and this corresponds to hypothesizing a linear growth shape. This is demonstrated in the model below, which assumes linear growth across the four equally-spaced measurement times:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} \qquad (6)$$

In the model above, fixing the first loading at zero determines that the estimated mean and variance of the level should be interpreted as the mean and variance of variable at the first measurement time. In this case, the level factor is an intercept. Other coding schemes can be used to hypothesize the same linear growth shape, resulting in identical fit statistics and estimated variance of the shape, but with different estimates of the mean and variance of the level, mean of the shape, and unique variances (Biesanz et al., 2004). The interpretation of the parameters of the level (i.e. mean and variance) will also change, because the level would not be interpreted as an intercept. This is demonstrated in the model bellow, which is mathematically equivalent to the model presented above:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} \tag{7}$$

Latent growth models with different loading patterns will be equivalent only if the loading pattern of one model can be obtained by a linear transformation of the loadings of the other model (Biesanz et al., 2004). If this condition is satisfied, the parameter estimates that would be obtained if another pattern of loadings had been used can be determined analytically as a function of the pattern of loadings chosen (Biesanz et al., 2004):

$$\Phi^{*} = T^{-1}\Phi T^{-1'} \tag{8}$$

$$\mu_\eta = T^{-1}\mu_\eta \tag{9}$$

where an asterisk indicates the transformed model and no asterisk indicates the original model, Φ is the variance/covariance matrix of the level and slope, μ_η is a vector containing the estimated mean of the level and slope, T^{-1} is a transformation matrix defined by $T^{-1} = (\Lambda_y{*'}\Lambda_y{*})^{-1}\Lambda_y{*'}\Lambda_y$, and Λ^{*} is the factor loading matrix with the new loadings. This transformation allows a researcher to examine what the parameters would be if a different measurement occasion had been used as a reference, without having to re-run the analysis. Although models obtained through linear transformations of the loading pattern are equivalent, each loading pattern requires a different interpretation of the mean and variance of the level and shape factors.

A model with fixed loadings is nested within a model with free loadings. Therefore, a chi-square difference test can be performed to test the adequacy of the hypothesis established by the fixed loadings (Meredith & Tisak, 1990). A researcher may want to test a specific hypothesis about the growth shape, and if the model with fixed loadings has adequate fit, the researcher may want to perform a chi-square difference test between this constrained model and the model where the loadings are freely estimated, to verify whether the fixed loadings do not significantly reduce model fit. If there is no significant difference between the constrained and unconstrained model, the researcher finds support for the model hypothesized by the fixed loadings. This chi-square test will be described in more detail later in the section about invariance.

The measurement part of univariate LGM provides information about whether the variance/covariance matrix of the observed variable across measurement occasions fits the growth trajectory specified by the model. The structural part of the model complements the measurement model by providing information about the means, variances and covariance of the level and shape factors.

In order to estimate the means of the level and shape factors, they are regressed on a scalar equal to one. The means $\mu_{\eta_{1i}}$ and $\mu_{\eta_{2i}}$ of the level and shape factors, respectively, are the path coefficients from the unit indicator to these factors. Because the level and shape factors are regressed on a constant, their variances become residual variances represented by the disturbances $\zeta_{\eta_{1i}}$ and $\zeta_{\eta_{2i}}$, respectively. The covariance between these disturbances $\sigma_{\zeta_{\eta_{1i}}\zeta_{\eta_{2i}}}$ is the covariance between the level and shape factors.

The structural part of a univariate LGM without predictors is:

$$\eta = \mu + \zeta \tag{10}$$

where the term μ is a 2 x 1 vector containing the means $\mu_{\alpha_i}, \mu_{\beta_i}$ of the level and shape factors, respectively. The mean of the level, μ_{α_i}, represents the average status of the observed variable at the reference occasion of measurement. The mean shape, μ_{β_i}, represents an average trajectory of growth across all the population. The term ζ represents a 2 x 1 vector of disturbances. The disturbances ζ_{α_i} and ζ_{β_i} are the deviations of the individual level α_i and shape β_i parameters from their population means μ_{α_i} and μ_{β_i} respectively.

In matrix format, this model is:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_{\alpha_i} \\ \mu_{\beta_i} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha_i} \\ \zeta_{\beta_i} \end{bmatrix} \tag{11}$$

The matrix Ψ is the variance/covariance matrix between the disturbances:

$$\Psi = \begin{bmatrix} \sigma^2_{\zeta_{\alpha_i}} & \sigma_{\zeta_{\alpha_i}\zeta_{\beta_i}} \\ \sigma_{\zeta_{\alpha_i}\zeta_{\beta_i}} & \sigma^2_{\zeta_{\beta_i}} \end{bmatrix} \tag{12}$$

where $\sigma^2_{\zeta_{\alpha_i}}$ is the variance of the level factor, $\sigma^2_{\zeta_{\beta_i}}$ is the variance of the shape factor, and $\sigma_{\zeta_{\alpha_i}\zeta_{\beta_i}}$ is the covariance between level and shape. The variances of the disturbances are equal to the variances of the factors because this model is unconditional, which means that the only predictor is a constant added to allow the estimation of means. If predictors are added to the model, the variances of these disturbances becomes residual variances, which represent the variability remaining in the level and shape factor scores after removing the effect of the predictor.

There are five important parameters estimated in the structural part of the univariate LGM: the means and variances of the level and shape factors, and their covariance. The mean of the level indicates the average of the observed variable at the measurement time chosen as reference. The variance of the level reveals whether individuals differ from each other on the observed variable at the reference time of measurement. The mean of the shape indicates the average growth trajectory. If growth is linear, the mean of the shape is equivalent to the growth rate of the individuals on the observed variable. A positive mean shape indicates that individuals increase in the observed variable with time, while a negative shape indicates a decrease in the observed variable. The variance of the shape provides information on whether individuals differ on growth trajectory. If the variances of the level and shape are statistically significant, then it is interesting to attempt to account for this variability by including predictors in the model (Willett & Keiley, 2000). The covariance between level and shape indicates whether there is a relationship between the status of the individuals on the outcome variable at the reference time of measurement and their growth trajectory.

The measurement and structural parts of the univariate LGM describe growth at the individual and group level, respectively. The same models can be expressed in terms of describing the variance/covariance matrix between observations across individuals across time and the vector of means of the observed variable at each measurement time (Biesanz et al., 2004). These two different formulations of the latent growth model are equivalent (MacCallum et al., 1997; Singer & Willett, 2003; Willett & Keiley, 2000; Willett & Sayer, 1994, 1996). The latent growth model in terms of describing the observed variance/covariance matrix and vector of means is (Biesanz et al., 2004; Bollen, 1989; MacCallum et al., 1997):

$$\Sigma_{yy} = \Lambda_y \Phi \Lambda_y' + \Theta_\varepsilon \quad (13)$$

$$\mu_y = \Lambda_y \mu \quad (14)$$

where Σ_{yy} is the *j* x *j* variance/covariance matrix of the observed variable measured in *j* occasions, Λ_y is a *j* x 2 matrix of factor loadings, Φ is the 2 x 2 variance/covariance matrix of the level and shape factors, Θ_ε is a *j* x *j* variance/covariance matrix of errors of measurement, and μ_y is a column vector of *j* observed means, and μ is a 2 x 1 vector containing the means μ_{α_i} and μ_{β_i} of the level and shape factors, respectively.

Conditional latent growth models

The LGM model presented above was unconditional in the sense that no predictors of growth were included in the model. However, if there is individual variability in the level and shape of growth, predictors of this variability can be sought. In LGM, two types of predictors can be included in the model: time-invariant and time-varying (Willett & Keiley, 2000). Time-invariant predictors are measured only once, and their relationship with the outcome variable is evaluated at a single measurement occasion. Time-varying predictors are measured multiple times and this enables the investigation of how the growth in the predictor affects growth in the outcome variable. Time-varying predictors must be assessed in the same measurement occasions the outcome variable was measured. Figure 2 presents an example of a latent growth model with three measurement times and one time-invariant predictor.

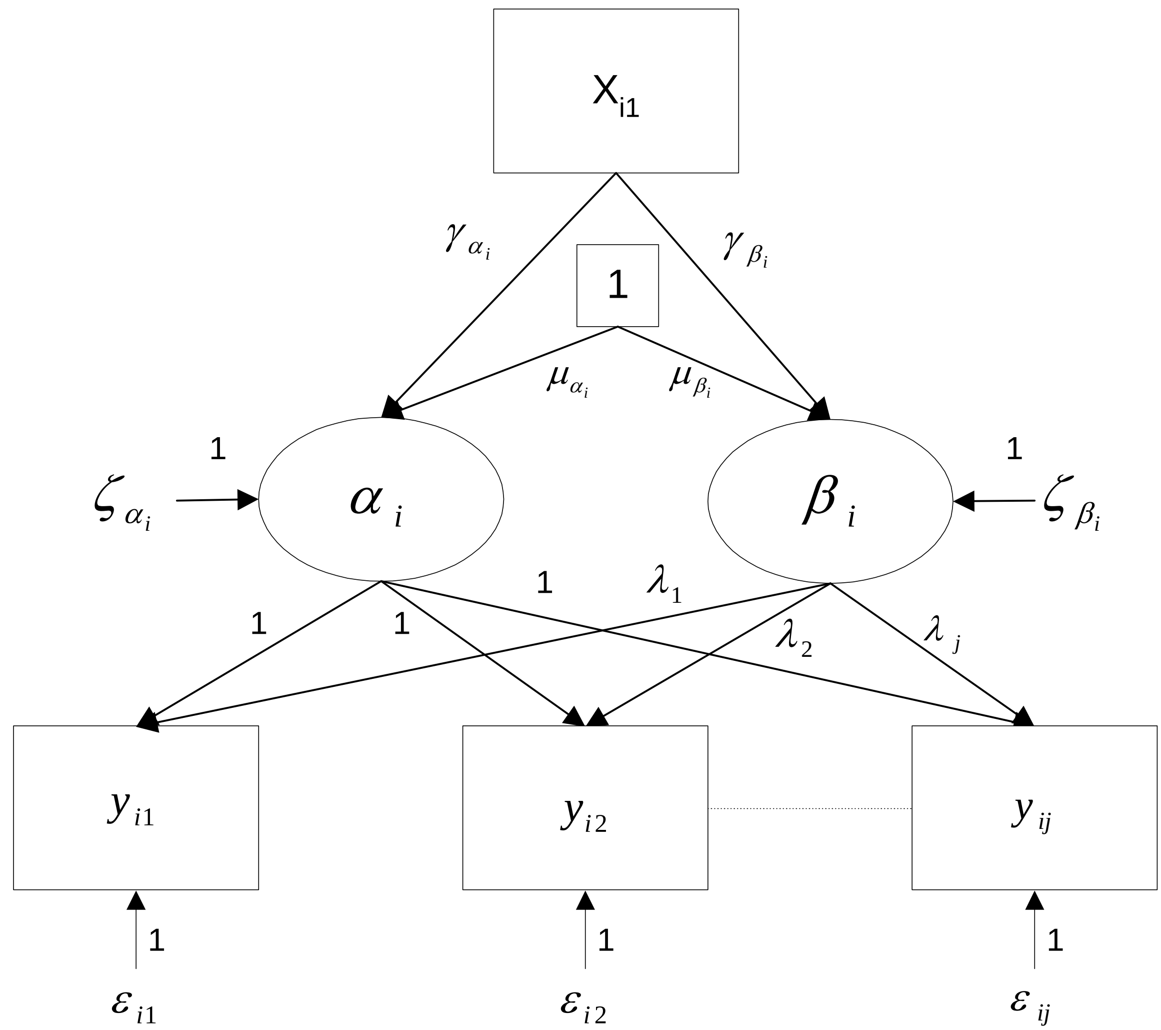

Figure 2. Latent growth model with one time-invariant predictor.

Predictors are inserted in the structural part of the LGM model, because the goal of the analysis is to evaluate the relationship of the predictors with the level and shape factors. LGM allows a lot of flexibility in the specification of predictors. For example, predictors may have their own measurement model, or may relate to the outcome variable through the mediation of a third variable. The structural part of a univariate LGM model where a single time-invariant predictor was included is presented below:

$$\eta = \mu + \Gamma\xi + \zeta \qquad (15)$$

where Γ is a 2 x 1 vector of regression coefficients relating the predictor to the level and shape factors, and ξ is the score on the predictor. The predictor can be either an observed variable or a latent variable.

The equation above in expanded format is:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_{\alpha_i} \\ \mu_{\beta_i} \end{bmatrix} + \begin{bmatrix} \gamma_{\alpha_i} \\ \gamma_{\beta_i} \end{bmatrix} [\xi] + \begin{bmatrix} \zeta_{\alpha_i} \\ \zeta_{\beta_i} \end{bmatrix} \qquad (16)$$

where γ_{α_i} is the regression coefficient between the predictor and the level parameter and γ_{β_i} is the regression coefficient between the predictor and the shape parameter.

In the conditional model, the interpretation of the parameters estimated in the structural model changes. The disturbances, ζ , become the residual variability not accounted for by the predictor. The matrix Ψ becomes the variance/covariance matrix of the level and shape parameters controlling for the relationship with the predictor (Willett & Keiley, 2000). More specifically, the variance of the level and shape parameters become residual variances after removing the part accounted for by the predictor. The covariance between level and shape becomes a partial covariance where the relationship with the predictor is removed.

In the structural part of the univariate LGM of one predictor, the two regression coefficients that indicate the relationship between the predictor and the level and shape parameters are estimated. The interpretation of these coefficients depends on whether the predictor is continuous or dichotomous. If the predictor is continuous, the coefficients indicate the average amount of change in the level and shape parameters that is expected with a unit change in the predictor. If the predictor is dichotomous, the coefficients are the expected difference in level and shape of the outcome between the two groups defined by the predictor variable.

Stoel, van den Wittenboer and Hox (2004) proposed that an alternative to include time-invariant covariates in an univariate LGM model is to regress the predictor directly on the observed indicators. They argue that when predictors are regressed on the level and shape factors, an assumption of full mediation is being made, which means that the direct effect of the predictor on the indicators is assumed to be zero. The authors recommend that when a researcher believes that there are direct effects of the predictor on the indicators, it is preferable to include predictors regressed directly on the observed indicators, or else the researcher may obtain poor model fit. Another advantage of this approach is that the effects of the predictors on the indicators do not need to follow the trajectory of growth hypothesized in the loadings of the shape factor. However, when the assumption of full mediation is tenable, regressing predictors on the level and shape parameters has the advantage of allowing an easy interpretation of the effects of the predictor on growth, and the calculation of the proportion of variance of the level and shape factors explained by the predictor. Stoel, van den Wittenboer and Hox (2004) model is illustrated in the figure below:

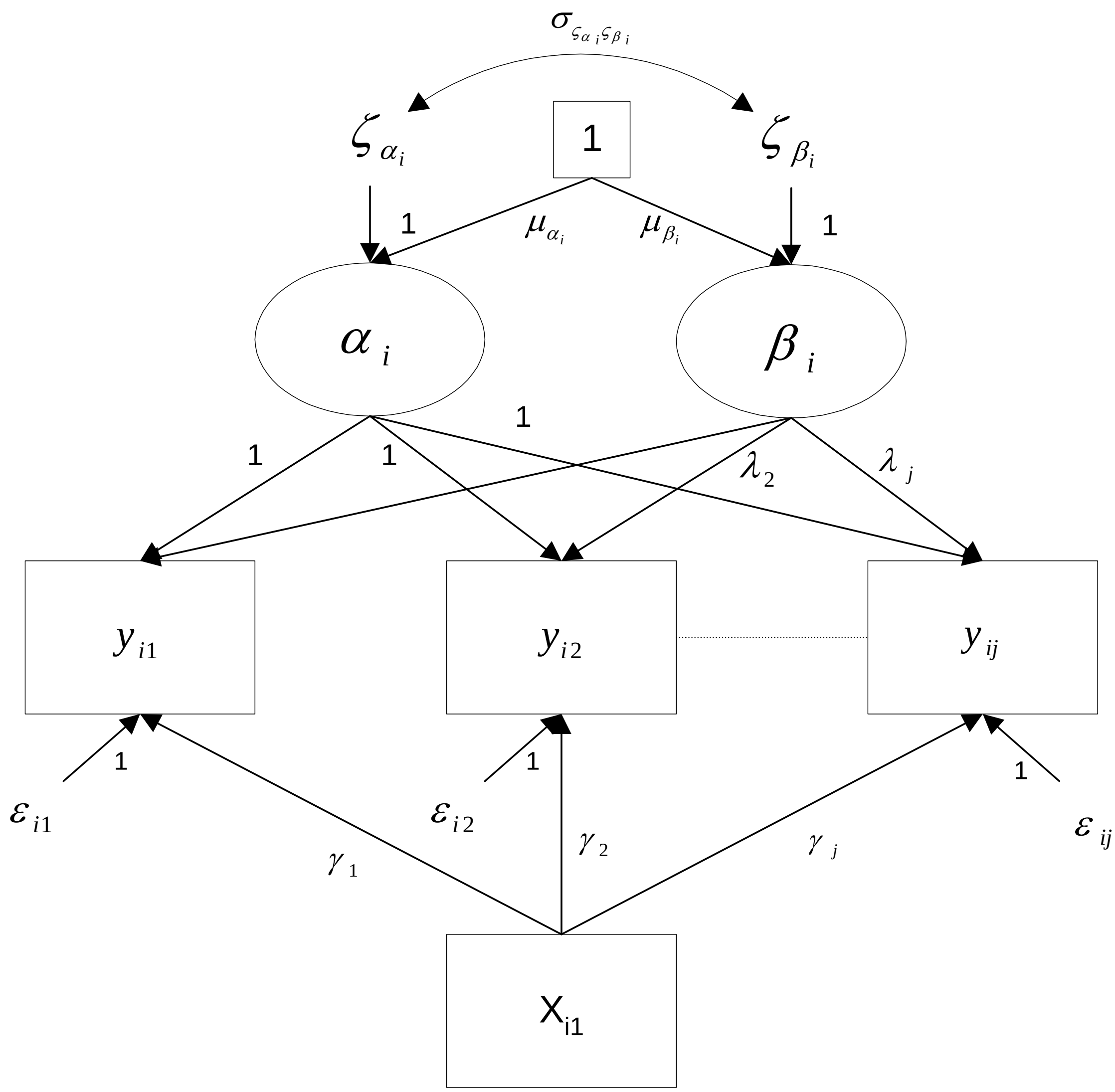

Figure 3. Alternative specification of latent growth model with one time-invariant predictor.

To include a time-varying predictor in the LGM model, it is necessary to establish a growth model for both the predictor and outcome variables. The measurement part of the LGM of the predictor is (Willett & Keiley, 2000):

$$x = \Lambda_x \xi + \delta \tag{17}$$

where x is a vector of observed scores on the predictor variable for individual i on times one to j, Λ_x is a j x 2 matrix containing a column of level coefficients and a column of shape coefficients for j measurement times, ξ is a 2 x 1 vector containing latent scores for the level factor α_{ξ_i} and shape factor β_{ξ_i}, and δ is a vector of errors in each of the j measurement occasions. In matrix format, this model is:

$$\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \\ \cdot \\ \cdot \\ \cdot \\ x_{ij} \end{bmatrix} = \begin{bmatrix} 1 & \lambda_{\xi 1} \\ 1 & \lambda_{\xi 2} \\ 1 & \lambda_{\xi 3} \\ 1 & \lambda_{\xi 4} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \lambda_{\xi\ j} \end{bmatrix} \begin{bmatrix} \alpha_{\xi_i} \\ \beta_{\xi_i} \end{bmatrix} + \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \delta_{i3} \\ \delta_{i4} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{ij} \end{bmatrix} \tag{18}$$

Once the measurement model of the predictor is established, the structural model defines the level and shape parameters of the predictor as explaining the level and shape parameters of the outcome. In other words, a common structural model states the hypothesized relationship between the predictor and the outcome. In the case of a time-varying predictor, both the level and shape parameters of the predictor are hypothesized to account for variability in the level and shape parameters of the outcome. The structural model is:

$$\eta = \mu + \Gamma \xi + \zeta \tag{19}$$

where Γ is a 2 x 2 matrix of regression coefficients relating the level and shape parameters of the predictor ξ to the level and shape parameters of the outcome η. In matrix format this model is:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_{\alpha_i} \\ \mu_{\beta_i} \end{bmatrix} + \begin{bmatrix} \gamma_{\alpha_i \alpha_{\xi_i}} & \gamma_{\alpha_i \beta_{\xi_i}} \\ \gamma_{\beta_i \alpha_{\xi_i}} & \gamma_{\beta_i \beta_{\xi_i}} \end{bmatrix} \begin{bmatrix} \alpha_{\xi i} \\ \beta_{\xi i} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha_i} \\ \zeta_{\beta_i} \end{bmatrix} \tag{20}$$

where $\gamma_{\alpha_i \alpha_{\xi_i}}$, $\gamma_{\beta_i \alpha_{\xi_i}}$, $\gamma_{\alpha_i \beta_{\xi_i}}$, $\gamma_{\beta_i \beta_{\xi_i}}$ are the regression coefficients between the level of the predictor and the level of the outcome, the level of the predictor and the shape of the outcome, the shape of the predictor and the level of the outcome, and the shape of the predictor and the shape of the outcome, respectively. The conditional latent growth model with one time-varying predictor is illustrated in figure 4.

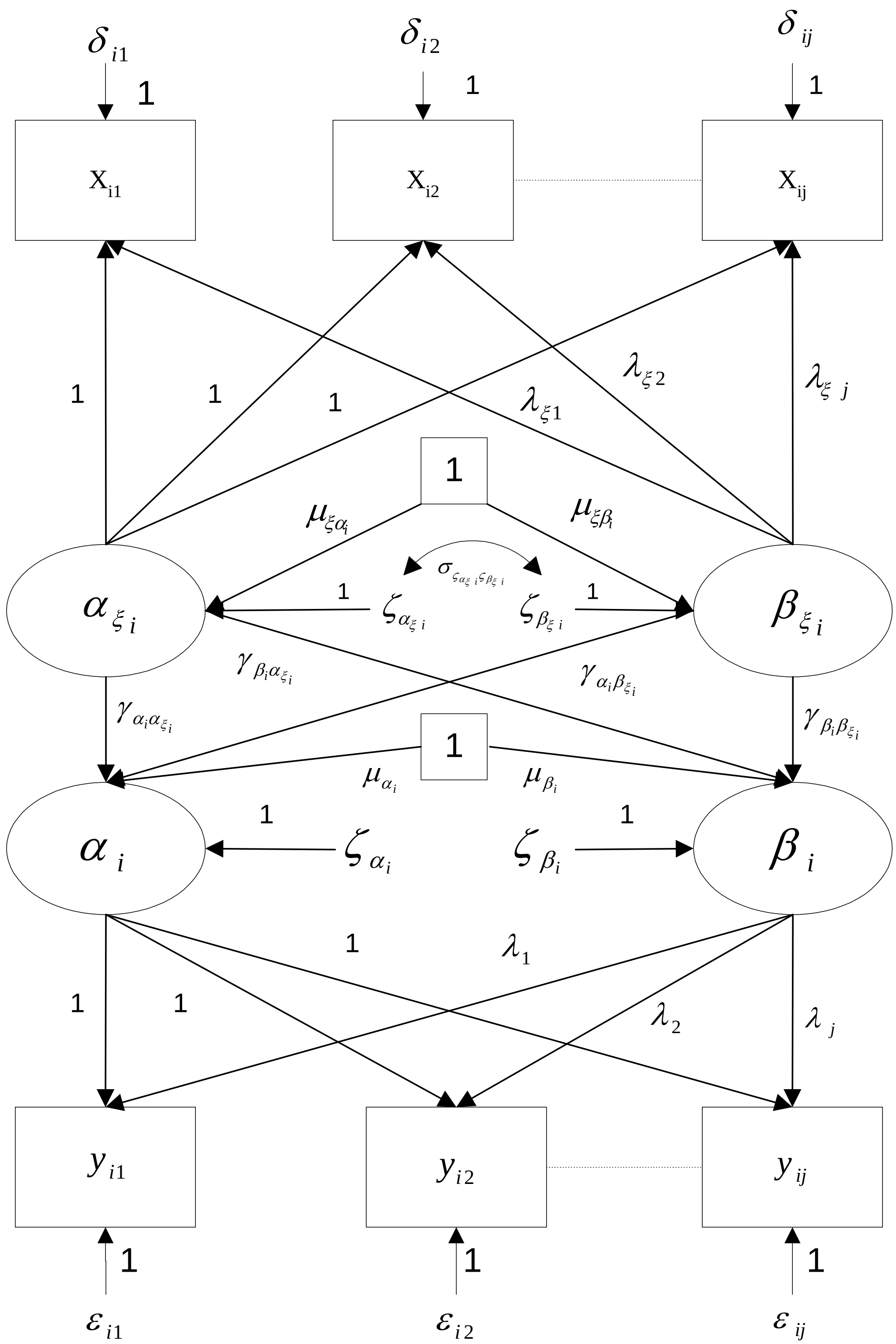

Figure 4. Latent growth model with one time-varying predictor

The coefficients in the Γ matrix are interesting theoretically to interpret, because they summarize the longitudinal relationship between the level and shape factors of the observed variable and the predictor. The $\gamma_{\alpha_i \alpha_{\xi_i}}$ coefficient indicates the extent that the level of the predictor affects the level of the outcome. If the first time of measurement is the reference, this parameter reveals whether the initial status of the outcome variable can be explained by the initial status of the predictor. The $\gamma_{\beta_i \alpha_{\xi_i}}$ coefficient indicates whether the level of the predictor explains the growth shape of the outcome. For example, in the case of a linear model where the first measurement time is the reference, a positive $\gamma_{\beta_i \alpha_{\xi_i}}$ would indicate that higher initial status on the predictor corresponds to faster growth of the outcome. The $\gamma_{\alpha_i \beta_{\xi_i}}$ coefficient reveals the extent that the shape of the predictor affects the level of the outcome. For example, if the last measurement time is defined as the reference, a positive $\gamma_{\alpha_i \beta_{\xi_i}}$ would indicate that individuals who growth faster on the predictor have a higher status on the outcome at the last measurement time. Finally, the $\gamma_{\beta_i \beta_{\xi_i}}$ coefficient indicates the effect of the shape of the predictor on the shape of the outcome. For example, in a linear model where this coefficient is positive, individuals with a higher growth rate in the predictor would be expected to have higher growth rate in the outcome.

In the conditional model with a time-varying predictor, the estimated variances of the level and shape are residual variances after the relationship with the predictor is accounted for, and the partial covariance controls for the predictor. This interpretation of the residual variance/covariance matrix in the conditional model remains the same if more time-invariant or time-varying predictors are included in the model. In a model with more predictors, each regression coefficient should be interpreted taking into account whether the coefficient is statistically significant, its sign, and the scale of measurement of the predictor.

Relationship between latent growth models and hierarchical linear models

The latent growth models presented in this study are within the structural equation modeling framework. Growth modeling can also be performed in the hierarchical linear modeling (HLM) framework, which consists in viewing repeated observations as nested within individuals (Raudenbush & Bryk, 2002). In the HLM framework, the level 1 (i.e.within-person) model represents the growth of each individual, and it is equivalent to the measurement part of the univariate LGM model because it allows the evaluation of whether the shape of growth implied by the model matches well the individual growth trajectories (Singer & Willett, 2003; Willett & Keiley, 2000; Willett & Sayer, 1994). The level 1 unconditional model for linear growth is presented in the equation bellow, using the notation defined by Raudenbush and Bryk (2002):

$$y_{ji} = \pi_{0i} + \pi_{1i} a_{ji} + e_{ji} \quad (21)$$

Where the observed score for person *i* at time *j* is a function of the person's initial status π_{0i}, growth rate π_{1i}, time of measurement a_{ji}, and error e_{ji}. The parameters

π_{0i} and π_{1i} correspond to the individual's level and shape, respectively, with the SEM framework.

The level 2 model represents the mean and variability in initial status and growth rate across individuals. It is equivalent to the structural part of the univariate LGM model (Singer & Willett, 2003; Willett & Keiley, 2000; Willett & Sayer, 1994). The level 2 unconditional model is presented in equations 21 and 23.

$$\pi_{0i} = \beta_{00} + r_{0i} \tag{22}$$
$$\pi_{1i} = \beta_{10} + r_{1i} \tag{23}$$

Where β_{00} are the mean intercept (i.e. level) and β_{10} is the mean growth rate (i.e. shape). The variances of the random effects r_{0i} and r_{1i} are the variances of the level and shape respectively, and the covariance between r_{0i} and r_{1i} is the covariance between level and shape.

In modeling growth with HLM, time-varying predictors are included in the level 1 model, while time-invariant predictors are included in the level 2 model. A model with one time-varying predictor and one time-invariant predictor is shown bellow.

Level 1 model:

$$y_{ji} = \pi_{0i} + \pi_{1i}a_{1ji} + \pi_{2i}a_{2ji} + e_{ji} \tag{24}$$

Where a_{1ji} is the time of measurement, and a_{2ji} is a time-varying predictor, which has a regression coefficient equal to π_{2i}.

Level 2 model:

$$\pi_{0i} = \beta_{00} + \beta_{01} X_{1i} + r_{0i} \quad (25)$$
$$\pi_{1i} = \beta_{10} + \beta_{11} X_{1i} + r_{1i} \quad (26)$$

Where X_{1i} is a time-invariant predictor, which has a regression coefficient of β_{01} on the individual level and a regression coefficient of β_{11} on the individual shape.

In the growth modeling context, Raudenbush and Bryrk (2002) argue that the dissimilarities between HLM and SEM are a matter of limitations of current software rather than real differences between the models. For example, when there are correlated errors, Raudenbush and Bryrk (2002) recommend using SEM software because they allow easy specification and estimation of correlations between errors. Little, Schnabel, and Baumert (2000) point out that, although it is possible to model correlated measurement errors within the HLM framework, it would require a difficult setup.

Assumptions and limitations of latent growth models

The latent growth models presented in this study assume an equal number and spacing of assessments for all individuals and that there is no missing data (Duncan, T. E. et al., 1999). Some researchers argue that LGM analysis must always involve these assumptions and that they are serious limitations of this method (e.g.Willett & Sayer, 1994). However, MacCallum et al. (1997) argue that these limitations are no longer true because of the development of full-information maximum likelihood estimation and software that can implement it. In the simple situation where individuals who were tested in different measurement times (or that have the same missing data pattern) can be grouped into a small number of sets, multiple-group latent growth modeling can be used to accommodate these irregularities. If irregularities in the measurement times and missing data patterns are complex, estimation of the LGM model can still be accomplished by using full-information maximum likelihood, which defines the likelihood function in terms of individual scores on observed variables, instead of the variance/covariance matrix (MacCallum et al., 1997).

Analyzing latent growth models requires some of the same assumptions of structural equation models, which can be relaxed if different estimation methods are used. For example, if maximum likelihood estimation is used, there is an assumption that the outcome variables are measured on a continuous scale and are multivariate normally distributed (Byrne & Crombie, 2003). However, if the outcome variables are categorical and non-normally distributed, other estimators can be used such as weighted least squares with corrected means and variances (WLSMV) (Muthén, L. K. & Muthén, 1998) or the Satorra-Bentler robust maximum likelihood estimator (Bentler, 1995).

The most serious assumption of LGM, which cannot be relaxed with current methods, is that it assumes that all individuals have the same functional form of development (e.g. linear, quadratic) (Hertzog & Nesselroade, 2003; Lawrence & Hancock, 1998).The loadings of the shape factor determine the functional form of development, which is also known as the basis function (Hertzog & Nesselroade, 2003) of development. In LGM, individuals can vary in the amount of change, as long as the form of development specified by the loadings of the shape factor fits all individuals. The assumption of a single basis function for all individuals can be violated if the dataset contains individuals from two populations with different forms of development. For example, a longitudinal dataset about the effects of a medicine for a terminal disease (e.g. cancer) may contain a group of individuals for whom the use of the medicine causes improvement with time, and another for whom the use of the medicine has no effect, and they present continuous health decline with time. If the two different groups can be identified, multi-group analysis can be performed specifying a different form of development for each group.

The LGM method presented above is univariate, because it modeled a single observed variable measured across multiple times. However, if the researcher is interested in studying the growth of several variables that jointly assess a latent construct (e.g. several test items measuring a single trait), aggregating the individuals' scores of these variables into composites of equally weighted items (e.g. using the mean or sum of the scores) and using univariate LGM to analyze the growth of the composite scores may fail to provide parameter estimates that correctly represent the growth of the latent variable. The reason is that univariate LGM of composites of multiple items treats the composites as if they were latent variables. In other words, it is assumed that all the variance in the composite is common variance from the multiple items, but the composites' variance may also contain the items' error variance and specific variance (Lomax, 1986), which are not related to the latent construct of interest. Under the factor analysis perspective, error variance is due to random measurement error and specific variance is due to systematic and reliable factors that are not part of the latent construct. The error variance and specific variance combined form the unique variance of an item (Lomax, 1986).

Measurement error and systematic factors may affect the variance/covariance matrix of the composites, leading latent growth analyses of these composites to yield biased parameter estimates. First, the variances of the composites measured at multiple times may be larger than the variances of the corresponding latent construct, because the total variance of a composite variable may contain not only the common variance of the items but also their unique variance. Second, with respect to the covariances between the composite variables, Fan (2003b) found, through a simulation study, that correlation coefficients between composites uncorrected for the items' measurement error are systematically biased downward. He also found that this bias is inversely related to the composites' reliability: the lower the reliability, the higher the bias will be.

Considering that biased parameter estimates may invalidate conclusions obtained from statistical analyzes, it is very important to choose a LGM method that effectively addresses the measurement error in analyses of latent variables measured by multiple indicators. The review of the LGM literature indicates that alternatives to perform LGM of multiple indicators measured with error have never been compared. In this dissertation, four different alternatives will be compared. These alternatives are to use univariate LGM to analyze the unadjusted variance/covariance matrix between composites, to correct the variance/covariance matrix for attenuation before fitting the univariate LGM, to estimate the error variance of the composite at each measurement time and perform univariate LGM with fixed error variances, and to use the curve-of-factors model. Each of these methods will be explained in detail in the next section.

Latent growth models of multiple indicators

The first alternative is to perform a univariate LGM analysis of a composite variable formed by equally weighted indicators of the latent variable (e.g. mean, sum). The review of the LGM applied literature indicates that this approach has been researchers' most common choice. For example, Mason (2001) used LGM to analyze the growth of self-esteem in adolescents and its relationship with delinquency. In his study, the self-esteem variable modeled was the average of individuals' scores on 10 items pooled from two different self-esteem scales. In another study, Chan et al. (2000) used LGM to investigate children's social skills, and obtained the outcome variable by summing the scores of 38 items of the parent version of the scale and 30 items of the teacher version of the scale. In the path diagram shown in figure 3, composites $c_1 \ldots c_j$ are the average or sum of the scores of each individual in *k* items at measurement times 1…*j*.

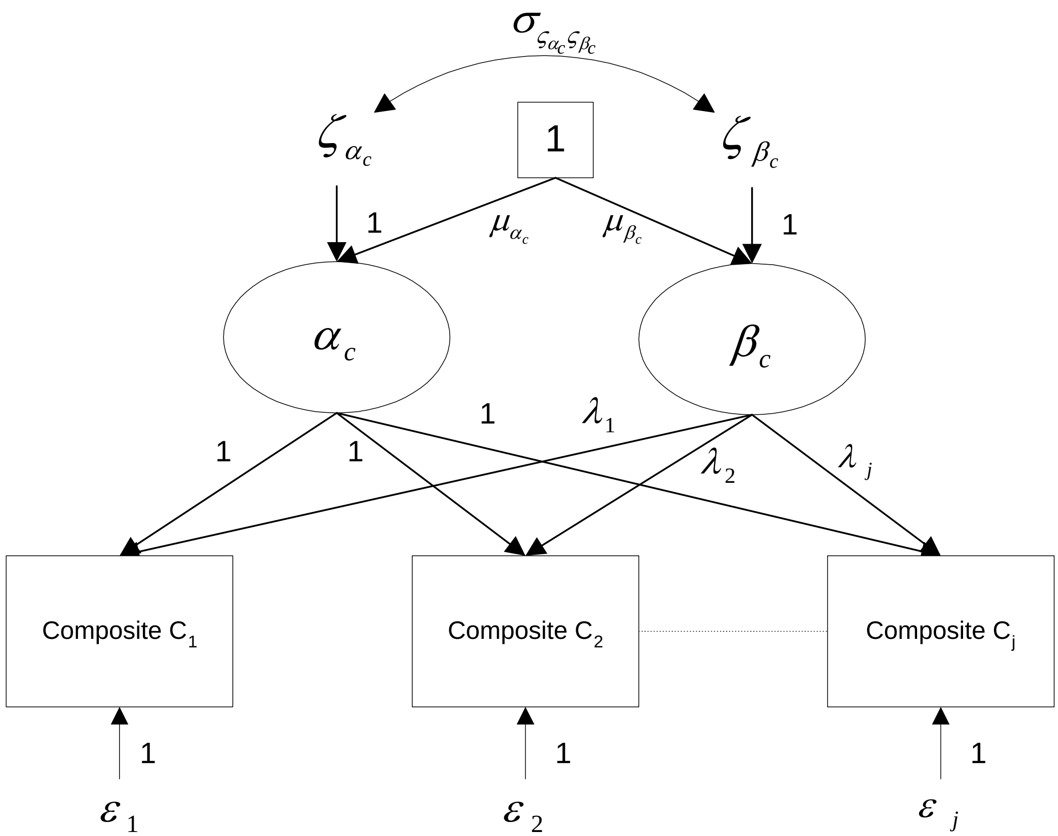

Figure 5. Univariate LGM of composites of multiple indicators

Bollen and Lennox (1991) demonstrated how the individuals' scores on a composite variable can differ from their scores on the latent variable, depending on the magnitude of the factor loadings of the indicators on the latent variable and the amount of measurement error: Consider a model with four indicators y_1, y_2, y_3, y_4, which are related to a single latent variable η through the factor loadings $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. The measurement errors of the indicators are $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$. If a researcher uses the sum c_1 of the indicators to model growth, this implies that:

$$c_1 = y_1 + y_2 + y_3 + y_4 \tag{27}$$

$$given \quad that \quad y_i = \lambda_i \eta + \varepsilon_i \quad \textbf{(} from \quad factor \quad analysis \textbf{)}$$

$$then \quad c_1 = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)\eta + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4)$$

$$defining \quad \lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \quad and \quad \varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4$$

$$then \quad c_1 = \lambda\eta + \varepsilon \tag{28}$$

This shows that the composite c_1 will differ from the latent variable η, unless the sum of the factor loadings is equal to one and there is no measurement error. The reliability $\rho_{C_j C_j^{'}}$ of a composite is related to the factor loadings, the variance of the measurement errors, and the number of items in the composite through the formula (DeShon, 1998; Reuterberg & Gustafsson, 1992):

$$\rho_{C_j C_j^{'}} = \frac{\left(\sum_{k=1}^{K} \lambda_k\right)^2}{\left(\sum_{k=1}^{K} \lambda_k\right)^2 + \sum_{k=1}^{K} \sigma_{\varepsilon_k}^2} \tag{29}$$

where λ_k is the factor loading of indicator k on a latent factor with K indicators, and $\sigma_{\varepsilon_k}^2$ is the error variance of indicator k.

The formula above indicates that the composite will be equivalent to the latent variable only if it has perfect reliability and that the reliability of the composite increases as the number of indicators increases. Researchers who perform univariate LGM of composite variables typically report a high estimate of the instrument's reliability, which may indicate that the measurement error is negligible and this approach yields acceptable results. Among the applied studies using LGM that were reviewed, the coefficient alpha estimates of the scales' reliability were typically between 0.7 and 0.9.

A second alternative to address measurement error in LGM is to correct the input variance/covariance matrix of the composites for attenuation (i.e. unreliability) before model fitting. A correction for attenuation estimates the correlation between the true scores of two variables. Because observed scores are, according to classical test theory, the sum of true scores and measurement error, the correction for attenuation estimates the correlation removing the effect of measurement error. In other words, the correction for attenuation estimates what the relationship between two variables would be if they were measured with perfect reliability (Muchinsky, 1996). The following formula for the correction for attenuation was proposed by Spearman (1904), and it has been frequently used in test construction (Crocker & Algina, 1986; Fan, 2003b; Lomax, 1986; Muchinsky, 1996):

$$\rho_{txty} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'}\rho_{yy'}}} \tag{30}$$

where ρ_{txty} is the correlation between true scores of two variables, ρ_{xy} is the observed correlation coefficient between variables x and y, and $\rho_{xx'}$, $\rho_{yy'}$ are the reliability estimates of x and y, respectively.

Fan (2003b) demonstrated that correcting the correlation between composite variables using the correction for attenuation formula with Cronbach's alpha provides similar results to using confirmatory factor analysis to model the correlation between the underlying latent variables. Through a Monte Carlo simulation study, Fan compared the correlation corrected for attenuation between two composites of multiple items and the correlation between the latent factors measured by the same items, estimated with a confirmatory factor analysis model. He simulated conditions with either four or eight items per factor, five inter-item correlations (.81, .64, .49, .36, and .25), two inter-factor correlations (0.4 and 0.6), and four sample sizes (50,100,200, and 400). The reliabilities of the composites resulting from the specification of the inter-item correlations were .57, .69, .79, .88, and .94 for the four-item composites and .72, .82, .88, .93, .97 for the eight-item composites. Fan found that the correlation between uncorrected composites was systematically biased downward, and that the bias increases as the reliability of the composites decreases. On the other hand, the estimates obtained with both the correction for attenuation and confirmatory factor analysis were unbiased. Furthermore, the standard errors of these inter-factor correlation estimates decreased as the sample size increased. He also found out that the standard error of the correlation between composites uncorrected for measurement error does not change with reliability. However, with the correction for attenuation and confirmatory factor analysis, increases in the reliability of the composites correspond to decreases in the standard error of the correlation between factors, resulting in more power to detect this correlation. There were no differences between the conditions with four items per factor and eight items per factor. He points out that the major advantage of using the correction for attenuation formula is its simplicity of implementation and recommends that the correction of attenuation should be performed whenever reliability coefficients are available.

There is a certain consensus in the literature that the type of reliability estimate employed in corrections for attenuation should be an internal consistency coefficient (Muchinsky, 1996; Osburn, 2000) because it assesses the degree of interrelatedness among the items administered in a given testing occasion (Cortina, 1993), and addresses the fact that the items used to form the composite are only a sample of possible items that measure the construct (Muchinsky, 1996). Using test-retest reliabilities would not be logical because the objective of LGM is to assess change across time, and this change should not be considered a source of error.

Cronbach's alpha is the most used internal consistency reliability estimate, and also the one that has been studied the most. Cortina (1993) reviewed the Social Sciences Citation Index for citations of Cronbach's alpha from 1966 to 1990 and found that it had been cited about 60 times per year in 278 different journals. Because the majority of the articles reviewed in the applied literature using LGM reported a Cronbach's alpha reliability estimate, this study will implement the correction for attenuation using this reliability coefficient. The formula for the Cronbach's alpha is:

$$\alpha_{C_j C_j^{'}} = \frac{K}{K-1}\left(1 - \frac{\sum_{k=1}^{K} \hat{\sigma}_k^2}{\hat{\sigma}_c^2}\right) \quad (31)$$

where K is the number of indicators combined to form each composite, $\hat{\sigma}_k^2$ is the observed variance of each indicator *k*, and $\hat{\sigma}_c^2$ is the variance of the composite. Although the coefficient alpha is traditionally represented as α, in LGM α has been used to represent the shape factor. To avoid confusion between the reliability coefficient and the shape factor, the term $\alpha_{C_j C_j^{'}}$ will be used here to refer to Cronbach's alpha reliability estimate. Derivations of the formula of Cronbach's alpha can be found in Lord and Novick (1968) and Crocker and Algina (1986).

Although Cronbach's alpha has been extensively used, there have been several studies showing that it underestimates reliability if the items of a test are congeneric (Komaroff, 1997; Osburn, 2000; Raykov, 1997, 1998, 2001). Given the classical test theory model $x_{ik} = \tau_{ik} + e_{ik}$, where x_{ik} is an observed score on item k for individual i, τ_{ik} is the true score, and e_{ik} is measurement error, two items are congeneric if for the same individual they produce different true scores τ_{ik} and different error variances $\sigma^2_{e_{ik}}$ (Lord & Novick, 1968). Two items are tau-equivalent if they produce the same true scores for each individual but different error variances. The most restrictive condition is of parallel measures, which requires that two items produce the same true score and the same measurement error for each individual. For Cronbach's alpha to yield unbiased reliability estimates, the items need to be essentially tau-equivalent, which means that the error variances are allowed to differ, but the true scores can only differ by a constant c (i.e. $\tau_{i2} = c + \tau_{i2}$) (Lord & Novick, 1968; Raykov, 2001). Cronbach's alpha becomes more robust to violations of tau-equivalence as the number of items increases (Osburn, 2000).

In structural equation modeling terms, congeneric indicators have different factor loadings and error variances, tau-equivalent indicators have equal factor loadings but different error variances, and parallel indicators have equal factor loadings and error variances (Reuterberg & Gustafsson, 1992). In the case where the latent variable is an endogenous variable, it is necessary to fix one of the factor loadings to one in order to set the scale of the factor. In this case, tau-equivalence requires that all the factor loadings of the indicators are equal to one (Bollen, 1989; Millsap & Everson, 1991). An additional level of classification of items into essentially parallel, essentially tau-equivalent, and essentially congeneric is only needed in the SEM framework if the latent means are being modeled (Millsap & Everson, 1991).These three types of items differ from parallel, tau-equivalent, and congeneric items because they allow items to have different intercepts, and consequently are less restrictive (Millsap & Everson, 1991).

Other methods to estimate reliability coefficients have been proposed such as Coefficient Theta (Bentler, 1972), the Maximized Lambda4 (Calender & Osburn, 1977), and Maximal Reliability (Li, H., Rosenthal, & Rubin, 1996). Although these methods are able to estimate the reliability of congeneric items, they have the disadvantage of being considerably more difficult to compute than Cronbach's alpha, which complicates their use in applied settings.

In univariate LGM modeling, the formula proposed by Spearman (1904) could be applied to correct the correlations between the composites using a reliability estimate of each composite. However, unreliability in the indicators affects not only the correlation between the composites but also inflates the composite's variance. It is possible to correct the variance of the composites for attenuation. The derivation of this correction comes directly from the definition of reliability in classical test theory: The reliability of a test is a measure of the degree of true score variation relative to observed score variation (Lord & Novick, 1968, p. 61). This definition is presented in the equation bellow:

$$\rho_{C_jC_j'} = \frac{\sigma^2_{T_i}}{\sigma^2_{C_i}} \qquad (32)$$

Where $\rho_{C_jC_j'}$ is the reliability of the composite, $\sigma^2_{T_i}$ is the variance of the true scores and $\sigma^2_{C_i}$ is the total variance of the composite. Conceptually, equation 32 is identical to equation 29, but equation 29 is presented in terms of item factor loadings and error variances. By re-arranging the terms of equation 32, the formula for the variance of the composite corrected for attenuation is obtained:

$$\sigma^2_{T_i} = \rho_{C_jC_j'}(\sigma^2_{C_i}) \qquad (33)$$

After both the correlations and variances of the composites are corrected for attenuation, the correlation matrix would then be transformed to a variance/covariance matrix, the vector of composite means would be added, and the resulting moment matrix could be used in the LGM analysis. To transform the corrected correlation of composites *x* and *y* into a covariance, it should be multiplied by the square root of the corrected variances of *x* and *y*. The mean of the composite can be calculated by either taking the average of the total individual scores on the composite or by summing the means of each item. According to classical test theory, it is not necessary to introduce any correction to the means because the population means of the observed scores are equal to the means of the true scores (Crocker & Algina, 1986), and the means of the observed scores are estimates of their population means.

The correction for attenuation assumes that all the measurement error in the multiple indicators is random, which implies that there is no systematic error and, consequently, the correlation between measurement errors from different testing occasions is zero (Crocker & Algina, 1986). However, as shown previously, with longitudinal data the error term may contain unique factors representing systematic variation in the indicators across time. If this is the case, using the correction for attenuation when there is systematic error may result in poor model fit. Another limitation of the correction for attenuation formula is that it is possible to obtain corrected values of the correlation coefficient larger than 1.0, which are theoretically impossible (Muchinsky, 1996). However, in cases where the assumption of independent errors is tenable, the correction for attenuation employed with latent growth models of composites may be advantageous because it is very simple to implement.

A third possible method to control for measurement error in LGM is to use the internal consistency reliability of the composite (i.e. Cronbach's alpha) to estimate the error variance of the composite. The error variance is estimated through the multiplication of $(1-\rho_{C_jC_j^{'}})$, which is the proportion of the variance in the composite due to measurement error, by the total variance of the composite $\sigma_{C_j}^2$ (Bollen, 1989). This is shown in the formula bellow:

$$\sigma_{\varepsilon_j}^2 = (1-\rho_{C_jC_j^{'}})\sigma_{C_j}^2 \tag{34}$$

where $\sigma_{\varepsilon_i}^2$ is the estimated error variance of the composite at measurement time j, $\rho_{C_jC_j^{'}}$ is the reliability coefficient of the composite, and $\sigma_{C_j}^2$ is the variance of the composite.

Once the error variances of the composites at each measurement time are estimated, they can be used in the latent growth model by setting all the error variances of the model equal to the estimated error variances. Consequently, the error variances are not free parameters, but are previously estimated and fixed in the model. Instead of error variances, the model estimates disturbance variances ζ_j, which are the reliable variances of the composites. The factors η_j are the reliable part of the composites at each testing occasion. A path diagram of this method is shown in figure 4.

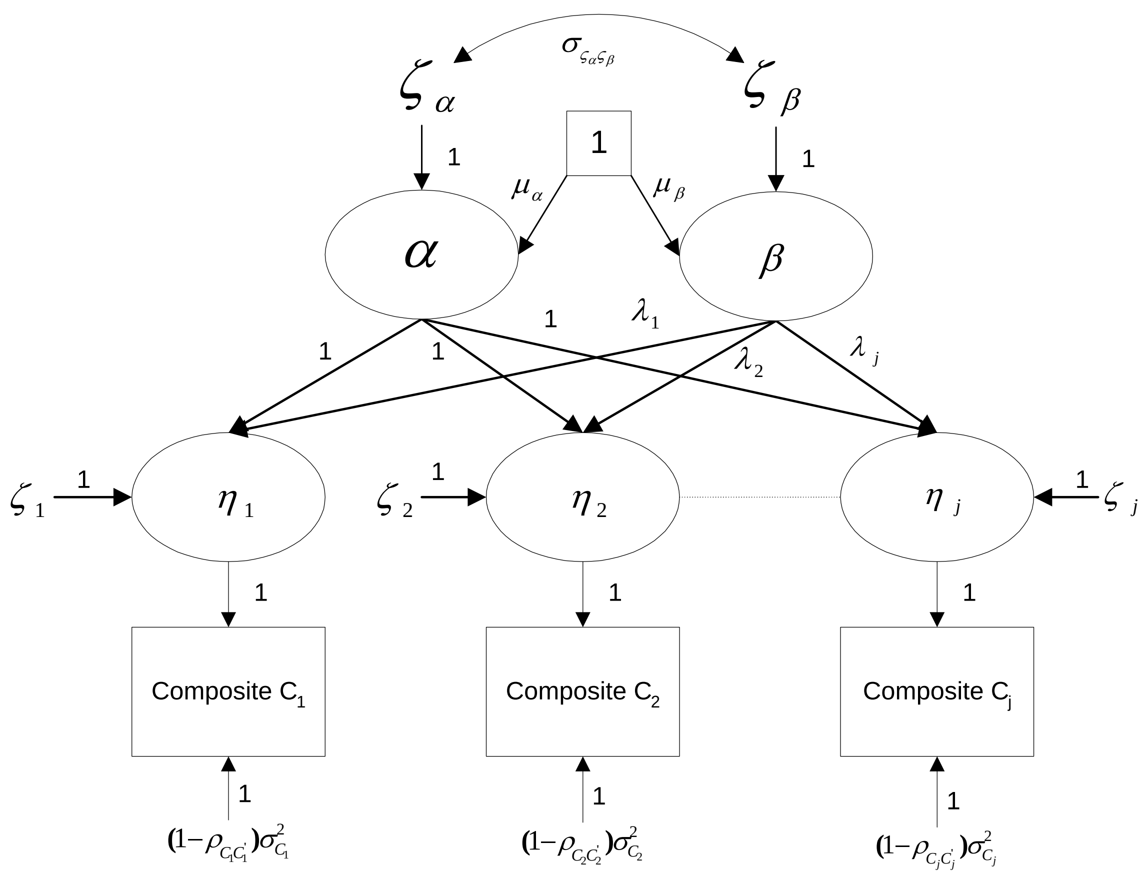

Figure 6. Univariate LGM with fixed error variances of the composites

This method has the advantage of being very easy to implement in LGM. It has been recommended in the manual of the LISREL 8 software (Jöreskog & Sörbom, 1996, p.196) and has been successfully used with structural equation modeling in applied studies. For example, Bandalos, Yates and Thorndike-Christ (1995) tested a model for statistics test anxiety where they used the reliability coefficient of the scales to estimate and fix the error variance of each of the latent variables. The same strategy to estimate the error variance of the latent variables was also employed by McWhirter, Hackett and Bandalos (1998) and Bandalos, Finney and Geske (2003).

A fourth way to control for measurement error is to estimate a measurement model for the latent variable together with the growth model. In LGM, this method can be implemented with the curves-of-factors model proposed by McArdle (1988) and Tisak and Meredith (1990). This is a second order latent growth model where the first order factors are latent variables measured multiple times, and the second order factors are the level and shape factors (Sayer & Cumsille, 2001). McArdle (1988) names this model the curve-of-factors model to contrast with the factor-of-curves model, which is also a second order LGM, but does not model the observed variables as indicators of a single latent variable. The factor-of-curves model (McArdle, 1988) is a second order model where an univariate LGM is specified for each indicator separately and then a common level and common shape are specified as second order factors. The level and shape factors of each indicator in the first order univariate LGM load on the second order common level and shape factors. The factor-of-curves model does not address the growth of a single latent contruct measured by multiple indicators, but the common growth of multiple observed variables. McArdle's terminology will be adopted in this study because it is more precise than the general term second order latent growth model.

Hancock et al. (2001) pointed out that the curve-of-factors model has the advantage of creating a theoretically error-free construct for growth modeling, instead of using composites of observed variables, which are contaminated by measurement error. The use of composites of indicators to model change in latent variables implies that the measurement error of each individual variable can be ignored. Furthermore, if the composite weights the variables equally, which is the case of the mean or sum of the variables, it is assumed that every indicator assesses the latent construct equally well (Sayer & Cumsille, 2001). The curve-of-factors model is presented in figure 2.

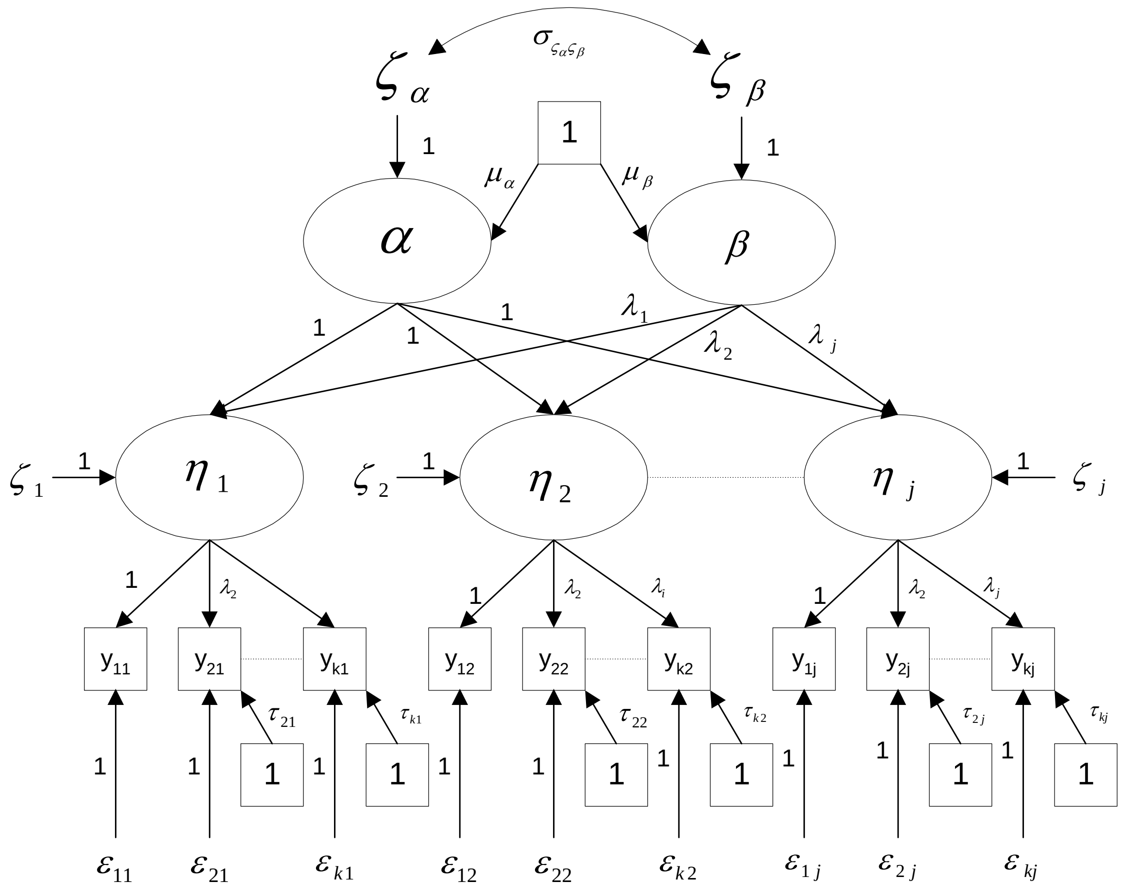

Figure 7. Curve-of-factors model

The curve-of-factors model is a combination of a common factor model and a latent growth model. The common factor part of the curve-of-factors model determines how well the indicators assess the latent variable in each time of measurement. The latent growth part determines the characteristics of the level and shape of the latent variable across time. The common factor part is (Bollen, 1989):

$$y = \tau + \Lambda \eta + \varepsilon \tag{35}$$

where y is a vector containing a set of raw scores for indicator k at each time j, τ is a vector of intercepts, Λ is a matrix of loadings, η is a vector of factor scores, and ε is a vector of random errors.

The estimation of a vector of intercepts, τ, is necessary because the expected value of y depends on the intercepts as shown below (Hancock et al., 2001):

$$\mathrm{E}[y] = \Lambda \mathrm{E}[\eta] + \tau \tag{36}$$

To estimate the vector of intercepts, τ, the indicators are regressed on a scalar equal to one (Hancock et al., 2001). In order to set the scale of measurement of the latent variables, it is necessary to select an indicator and fix its loading to one and its intercept to zero in each measurement occasion (Bollen, 1989; Hancock et al., 2001). This is exemplified in the model below, which has four indicators per factor and three measurement times:

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{12} \\ y_{22} \\ y_{32} \\ y_{42} \\ y_{13} \\ y_{23} \\ y_{33} \\ y_{43} \end{bmatrix} = \begin{bmatrix} 0 \\ \tau_{21} \\ \tau_{31} \\ \tau_{41} \\ 0 \\ \tau_{22} \\ \tau_{32} \\ \tau_{42} \\ 0 \\ \tau_{23} \\ \tau_{33} \\ \tau_{43} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ \lambda_{41} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{23} \\ 0 & 0 & \lambda_{33} \\ 0 & 0 & \lambda_{43} \end{bmatrix} \begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{41} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \\ \varepsilon_{42} \\ \varepsilon_{13} \\ \varepsilon_{23} \\ \varepsilon_{33} \\ \varepsilon_{43} \end{bmatrix}$$

The matrix Λ specifies the factor loadings of the indicators on the latent variable in each measurement time. The factor loadings of the indicators at a given measurement are equal to λ_{kj}, but are equal to zero at the other measurement times, because any correlation between indicators across measurement times is explained by either the latent growth part of the model or by unique factors included in the error term.

In the common factor part of the curve-of-factors model, the variance of the indicators is divided into common variance and unique variance. The unique variance contains error variance and specific variance. The curve-of-factors model can either assume no specific factors for each indicator, or imply the existence of specific factors by allowing the measurement errors of the indicators to correlate across testing times. The difference between modeling correlated measurement error in univariate LGM and in the curve-of-factors model is that in univariate LGM the errors of the composites may be allowed to correlate, while in the curve-of-factors model the errors of the indicators may be allowed to correlate. Correlated errors of composites have a different meaning of correlated errors of indicators. Correlated errors of composites represent the part of the *total variance* of the composites not accounted by the level and shape factors but that varies with time. Correlated errors of indicators represent the part of the indicator not accounted by the common factor (i.e. the specific factor) that varies meaningfully with time.

As an illustration of the flexibility of the curve-of-factors model with respect to the unique part of each indicator, the variance/covariance matrix of errors for a model with three indicators and two measurement times allowing heteroscedasticity and correlated errors across measurement times (but not within measurement times) is presented below:

$$\Theta_{\varepsilon} = \begin{bmatrix} \sigma^2_{\varepsilon_{11}} & 0 & 0 & \sigma_{\varepsilon_{11}\varepsilon_{12}} & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_{21}} & 0 & 0 & \sigma_{\varepsilon_{21}\varepsilon_{22}} & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_{31}} & 0 & 0 & \sigma_{\varepsilon_{31}\varepsilon_{32}} \\ \sigma_{\varepsilon_{11}\varepsilon_{12}} & 0 & 0 & \sigma^2_{\varepsilon_{12}} & 0 & 0 \\ 0 & \sigma_{\varepsilon_{21}\varepsilon_{22}} & 0 & 0 & \sigma^2_{\varepsilon_{22}} & 0 \\ 0 & 0 & \sigma_{\varepsilon_{31}\varepsilon_{32}} & 0 & 0 & \sigma^2_{\varepsilon_{32}} \end{bmatrix} \quad (37)$$

Once the common factor part of the indicators is specified, the growth of the latent variable uncontaminated by measurement error is addressed by the specification of the latent growth part of the curve-of-factors model, which includes the level and shape factors as second order factors. This part of the model is (Hancock et al., 2001):

$$\eta = \Gamma \xi + \zeta \quad (38)$$

where η is a vector of factor scores, Γ is a matrix of second-order factor loadings reflecting the growth pattern of the latent variable, ξ is a vector of latent scores capturing the level, α, and shape, β, parameters of the latent variable, and ζ is a vector of random normal disturbances.

In matrix form, the latent growth part of the curve-of-factors model is:

$$\begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \\ 1 & \lambda_3 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \zeta_{i1} \\ \zeta_{i2} \\ \zeta_{i3} \end{bmatrix} \quad (39)$$

This part of the curve-of-factors model is identical to a univariate LGM model, except that the outcome variables are latent variables, not observed variables. Therefore, the loadings of the latent growth part of the curve-of-factors model can be fixed to values that reflect a specific hypothesis about the shape of growth, or be free to estimate an unspecified growth curve. In the same way as undertaken for the univariate LGM, fixing one loading of the shape parameter equal to zero defines the corresponding measurement time as a reference for the interpretation of the latent means of the level and shape parameters. The model for these latent means, which corresponds to the structural part of univariate LGM, is:

$$\xi = \mu + \zeta \tag{40}$$

where ξ is a vector containing the level and shape parameters for each individual *i*, μ is the vector of latent means μ_α, μ_β of the level and shape, respectively, for all individuals, and ζ is a vector of disturbances.

In matrix format, this part of the curve-of-factors model is:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha_i} \\ \zeta_{\beta_i} \end{bmatrix} \tag{41}$$

The variances and the covariance of the disturbances are also estimated because they correspond to the variances and covariance of the level and shape factors. The matrix Ψ is the variance/covariance matrix between the disturbances:

$$\Psi = \begin{bmatrix} \sigma^2_{\zeta_\alpha} & \sigma_{\zeta_\alpha \zeta_\beta} \\ \sigma_{\zeta_\alpha \zeta_\beta} & \sigma^2_{\zeta_\beta} \end{bmatrix} \tag{42}$$

where $\sigma^2_{\zeta_\alpha}$ is the variance of the level factor, $\sigma^2_{\zeta_\beta}$ is the variance of the shape factor, and $\sigma_{\zeta_\alpha \zeta_\beta}$ is the covariance between level and shape.

In the curve-of-factors model, typically the parameters of interest are the first-order factor loadings λ_j, the mean μ_α of the level factor and its variance $\sigma^2_{\zeta_\alpha}$, the mean μ_β of the shape factor and its variance $\sigma^2_{\zeta_\beta}$, and the covariance $\sigma_{\zeta_\alpha\zeta_\beta}$ between the level and shape factor. If any of the second order factor loadings in the matrix Γ is also freely estimated, this coefficient also needs to be interpreted. Furthermore, the intercepts, τ_{jk}, of the observed variables are estimated, but usually they are not of theoretical interest (Hancock et al., 2001).

Even if measurement error is adequately accounted for by the measurement model at each testing time, parameter estimates may still be biased because of measurement non- invariance across testing times. Horn and McArdle (1992) and Meredith and Horn (2001) emphasize that measurement invariance is essential to scientific inference because if there is no evidence that one instrument applied to different conditions is measuring the same construct, any differences in group parameter estimates (e.g. means, variances, correlations) could be due to the fact that two different constructs were measured. The next section provides details about the issue of measurement invariance in structural equation modeling and latent growth modeling.

Measurement Invariance

Measurement invariance concerns whether the same instrument under different conditions (e.g. different groups of examinees, different occasions) yields measures of the same attribute (Horn & McArdle, 1992). In structural equation modeling, the term measurement invariance is used to refer to stability of parameters across populations and across conditions. For example, if the instrument is administered to two groups of elementary school students, one from low income families and the other from high income families, the instrument will be invariant if the item parameters are equivalent across groups. In the context of growth modeling, an instrument is invariant if the item parameters are constant regardless of whether the instrument was administered to students in fourth, fifth or sixth grade.

Horn, McArdle and Mason (1983) and later Horn and McArdle (1992) classified invariance conditions into metric invariance and configural invariance. They defined metric invariance as equality of factor loadings across two conditions. According to the authors, metric invariance provides a basis for inference that the same construct is being measured in different conditions. Configural invariance was defined as the stability of the configuration of zero and non-zero factor loadings (Horn & McArdle, 1992). This definition means that if an instrument has configural invariance, all the items that have non-zero loadings on a latent factor at one condition will also have non-zero loadings on the same factor in a different condition, but the actual value of the loadings may differ between conditions. This implies that instruments which only have configural invariance do not allow comparisons of constructs across conditions, because different values of factor loadings across conditions may indicate that the nature of the construct being measured has changed from one condition to the other.

Meredith (1993) developed a more detailed measurement invariance taxonomy, which defines four types of invariance. These types of invariance are progressively less restrictive. The most restrictive type is strict factorial invariance, which requires that the loadings, error variances, and intercepts are the same across multiple conditions. Strong factorial invariance, on the other hand, requires that the loadings and the intercepts are constant across conditions, but it does not require constant error variances. Next, weak factorial invariance only requires the loadings to be equivalent across conditions. The least restrictive condition is configural invariance, which expands the definition presented by Horn and McArdle (1992). Meredith (1993) defines configural invariance as a situation where not only the pattern of zero and non-zero loadings of the items is constant across conditions, but also the sign of the non-zero loadings is the same for all the conditions examined. This definition implies that the items of an instrument which have positive loadings on the latent construct at one condition (e.g. the examinees are sixth grade Hispanic students) must also have positive loadings on the construct at a different condition (e.g. the examinees are seventh grade white students), and none of the loadings that are non-zero in the first condition can become zero on the second condition, or vice versa.

Meredith's taxonomy considers that both the covariance and mean structure are being modeled. In structural equation modeling, sometimes only the covariance structure is being investigated, and invariance of intercepts is not required to be addressed.

Horn and McArdle's (1992) argue that metric invariance is necessary for a researcher to be able to infer that the same construct is being measured at two different conditions, which is equivalent to weak factorial invariance in Meredith's (1993) taxonomy. If this minimum requirement is not met, the measurement is considered to be non-invariant across conditions, and comparisons cannot be made. Non-invariance includes configural invariance, but also includes situations where loadings have different signs in each measurement condition or the pattern of zero/non-zero loadings changes.

Using univariate LGM to model composites of multiple indicators may result in both biased parameter estimates and poor model fit if there is non-invariance. This may happen because, in univariate LGM of composites, the loadings, error variances and intercepts of the items are not estimated, which corresponds to assuming that they are invariant across measurement occasions (i.e. there is strict factorial invariance). Poor model fit or biased parameter estimates could arise as a consequence of this assumption not being met.

To demonstrate why latent growth modeling of composites of multiple indicators could result in poor fit or biased parameter estimates, consider an univariate LGM model of a composite variable collected at four measurement times:

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \tag{43}$$

The composites are the sum of four indicators:

$$c_j = y_{1j} + y_{2j} + y_{3j} + y_{4j} \tag{44}$$

Given that $y_{kj} = \lambda_{kj}\eta + \varepsilon_{kj}$, from factor analysis, the univariate LGM model can be re-written as:

$$\begin{bmatrix} (\lambda_1+\lambda_2+\lambda_3+\lambda_4)\eta_1+(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4) \\ (\lambda_1+\lambda_2+\lambda_3+\lambda_4)\eta_2+(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4) \\ (\lambda_1+\lambda_2+\lambda_3+\lambda_4)\eta_3+(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4) \\ (\lambda_1+\lambda_2+\lambda_3+\lambda_4)\eta_4+(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} \quad (45)$$

Suppose that there is strict factorial invariance, which means that the factor loadings and errors remain constant across time. Suppose that the indicators' loadings are equal to 0.5 and their error variances are 0.25. Replacing the factor loadings and error variances in the previous equation:

$$\begin{bmatrix} 2\eta_1+1 \\ 2\eta_2+1 \\ 2\eta_3+1 \\ 2\eta_4+1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}$$

The matrix equation can be converted into a series of linear equations presenting the latent factors as a function of the level and shape scores:

$$\begin{aligned} \eta_1 &= 0.5(\alpha+\varepsilon_{i1}-1) \\ \eta_2 &= 0.5(\alpha+\beta+\varepsilon_{i2}-1) \\ \eta_3 &= 0.5(\alpha+2\beta+\varepsilon_{i3}-1) \\ \eta_4 &= 0.5(\alpha+3\beta+\varepsilon_{i4}-1) \end{aligned} \quad (46)$$

Because there is strict factorial invariance, the linear model hypothesized by the loadings of the shape factor (i.e. 0, 1, 2, 3) can be tested. The loadings of the indicators are a multiplicative constant and the error variances are an additive constant across measurement times, so they have no effect on the estimation of the level and shape factors.

However, suppose there is configural invariance. Suppose the loadings of the indicators at the first measurement and last measurement time are 0.5 and the error variances are 0.25, while the loadings at the second and third measurement times are 0.75 and the error variances are 0.5.

$$\begin{bmatrix} 2\eta_1 + 1 \\ 3\eta_2 + 2 \\ 3\eta_3 + 2 \\ 2\eta_4 + 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} \tag{47}$$

Solving for the latent factors as a function of the level and shape factors produces the following equations:

$$\begin{aligned} \eta_1 &= 0.5(\alpha + \varepsilon_{i1} - 1) \\ \eta_2 &= 0.33(\alpha + \beta + \varepsilon_{i2} - 2) \\ \eta_3 &= 0.33(\alpha + 2\beta + \varepsilon_{i3} - 2) \\ \eta_4 &= 0.5(\alpha + 3\beta + \varepsilon_{i4} - 1) \end{aligned} \tag{48}$$

In this situation, the loadings and error variances have different multiplicative and additive effects on the equations at each measurement time, creating the appearance of curvilinear growth. However, the hypothesis being tested, which is specified by the loadings of the shape factor, is that the growth is linear. Consequently, this model would fit the data poorly and the parameter estimates would be biased, although the differences between the latent factors $\eta_1 \ldots \eta_4$ could be truly linear.

The curve-of-factors model allows the examination of whether invariance holds across measurement times, because item level parameters are estimated. With the curve-of-factors model, the invariance assumption may be tested using only the measurement model, without specifying the second-order model. McArdle (1988) points out that the minimum requirement for the curve-of-factors model to meaningfully assess longitudinal change in the latent variable is that there is weak factorial invariance. If there is non-invariance, the estimates of the means and variances of the level and shape, and of the covariance between level and shape may be biased. This bias may happen as a consequence of the need to set the scale of the factors at each measurement time, which is done by fixing the loading of an indicator of each factor to one. Where there is non-invariance, bias may arise because fixing the loading of an indicator to one on each measurement time is equivalent to assuming that a unit change in the indicator fixed to one on the first measurement time means the same thing as a unit change in the indicator fixed to one on the other measurement time. This assumption will affect the estimates of the means and variances of the level and shape and their covariances, leading to biased estimates, because in reality the population loadings of the indicators whose loadings were fixed to one are different.

The test of invariance of the loadings (i.e. metric or weak factorial invariance) provides support for the hypothesis that the same indicators used at different situations measure the latent construct in the same way (Meredith & Horn, 2001). Testing for the invariance of both factor loadings and error variances is equivalent to testing for equal reliabilities across testing times (Raju, Laffitte, & Byrne, 2002). A researcher may test for all types of invariance in Meredith's taxonomy. There has been debate about whether a researcher should test strict factorial invariance first and if it does not hold, test less restrictive forms of invariance, or start with weak factorial invariance, then strong and last strict factorial invariance (Byrne, Shavelson, & Muthén, 1989; Raju et al., 2002).

Little (1997) argues that two different rationales can be used to test invariance: the statistical or the modeling rationales. A test for invariance according to the statistical rationale consists in placing equality constraints on each parameter hypothesized to be invariant across measurement occasions and comparing the fit of the constrained model with the fit of the unconstrained model. The constrained model is nested within the unconstrained model, and therefore a chi-square difference test can be used to evaluate the loss of fit introduced by the constraints (Sayer & Cumsille, 2001). This test is performed by taking the difference between the chi-square fit statistics of the constrained and unconstrained model (i.e. $\Delta\chi^2 = \chi^2_c - \chi^2_u$, where $\Delta\chi^2$ is the difference between the chi-squares, χ^2_c and χ^2_u are the chi-squares of the constrained and unconstrained model, respectively). The difference $\Delta\chi^2$ can be tested for significance with Δdf degrees of freedom, where Δdf is the difference between the degrees of freedom of the constrained model df_c and the unconstrained model df_u. The null hypothesis for this test is that there is invariance of the constrained parameters. If the researcher fails to reject the null hypothesis, it can be concluded that the constraints do not significantly reduce model fit, and the measurement invariance of the constrained parameters is supported. If the researcher rejects the null hypothesis, it means that the constraints over-simplify the model (Kline, 1998), and measurement invariance of the constrained parameters is not supported.

The modeling rationale for invariance testing consists in placing constraints in all parameters hypothesized to be invariant and evaluating the overall fit of the model. In this case, both global model fit indices (e.g. χ^2, NFI, CFI, TLI, RMSE) and indices of local misfit (e.g. Lagrange multiplier tests, residuals) should be inspected to establish a condition of invariance. The statistical rationale is more flexible than the modeling rationale, and it is especially useful to test a hypothesis about the invariance of a specific parameter. The modeling rationale is useful for large models with numerous constrained parameters, because the chi-square difference test is overly sensitive to model misfit with many parameters and large sample sizes (Little, 1997).

A serious limitation of current invariance tests is that because the loading of one indicator is usually fixed to one, this loading cannot be tested for invariance. Cheung and Rensvold (1999) proposed to perform the invariance test as many times as there are indicators, and switch the loading being fixed to one at each time. Invariance would only be assumed if none of these tests resulted in a significant chi-square difference test. However, Hancock, Stapleton, and Arnold-Berkovits (2004) argue that fixing the loading for an indicator to one not only assumes that this specific indicator is invariant across populations but also shifts the values of the other parameters in a compensatory way in order to result in the same observed variance-covariance matrix. They have demonstrated that tests of invariance will only yield adequate results if the assumption of invariance being made when fixing to one the loading of an indicator at each measurement time is adequate.

Although the issues presented above indicate a need for more research about tests of invariance, the focus of this dissertation is on consequences of different types of invariance on the parameter estimates and model fit of latent growth models of multiple indicators. In this dissertation, Meredith’s (1993) taxonomy will be used as a guide to create different invariance conditions and observe whether the methods of latent growth modeling can produce adequate estimates under these conditions.

Besides measurement error and measurement invariance in LGM of multiple-indicators, this dissertation will also address the sample size necessary for estimating latent growth models. Although there is a vast literature about sample size requirements for SEM, the sample size necessary for estimating latent growth models has received little attention. The goal of this dissertation is to provide some guidance about which sample size is large enough to ensure stable and unbiased parameter estimates, adequate fit indices, and adequate power, for each of the four LGM methods for multiple indicators. The section below presents some of the guidelines that have been provided to choose an adequate sample size in SEM, and reviews the few studies that have been published about sample size in LGM.

Sample-size necessary for fitting latent growth models

Parameter estimation in latent growth modeling, as well as in structural equation modeling in general, depends on sample size. The importance of sample size comes from the fact that the larger the sample size, the better the efficiency of parameter estimation. Efficiency refers to the size of the variance of the sampling distribution of the estimate. An estimator will be relatively more efficient than another if the variance of the sampling distribution of estimates is smaller (Wackerly, Mendenhall, & Scheaffer, 1996). A smaller variance of the sampling distribution corresponds to a smaller standard error, which in turn leads to increased power to detect significant effects. Furthermore, because parameter estimation in SEM is usually performed in an iterative process, a smaller variance of the sample statistics obtained with using a large sample size will result in better convergence rates. Also, estimates obtained with larger sample sizes will tend to cross-validate better.

With respect to the assessment of fit in SEM, increases in sample size correspond to increases in the chi-square statistic associated with the model. Sample size also affects fit indices by either influencing the sampling distribution of the fit index (e.g. NFI, GFI, AFGI) or by entering directly into the calculation of the fit index (e.g. TLI, IFI) (Bollen, 1990). The fit indices whose sampling distributions are influenced by sample size increase as sample size increases, while the ones that include sample size in their formula keep their expected value constant regardless of sample size (Bollen, 1990).

A variety of guidelines have been proposed to determine adequate sample sizes for estimating structural equation models, but these recommendations have not converged to a common agreement yet. The lack of agreement with respect to sample sizes is due to the fact that the necessary sample size depends on several different conditions, such as the complexity of the model, number of parameters being estimated, number of indicators per latent variable, distribution and reliability of the observed variables, strength of association between the indicators and the latent variables, estimation method, and amount of missing data (Hamilton, Gagne, & Hancock, 2003; Jackson, 2003; Muthén, L. K. & Muthén, 2002).

As a consequence of the many conditions that can affect the sample size necessary for estimating a structural equation model, several different criteria to determine minimum sample sizes have been recommended. One of these criteria is absolute sample size. For example, for maximum likelihood estimation of SEM models, a minimum of 150 (Anderson & Gerbing, 1988) or 200 (Jackson, 2003) observations has been recommended. However, other estimation methods such as the asymptotically distribution free (ADF) method may require more than 500 observations (Anderson & Gerbing, 1988). Another criterion for determining necessary sample size is based on the number of subjects (N) per parameter estimated in the model (q). For example, Kline (1998) recommends a minimum N:q ratio of 10:1 and ideally 20:1, while Bentler (1995) argues that with normally distributed data, a ratio of 5:1 could suffice. Jackson (2003) found that for a correctly specified model, an increase in the N:q ratio has a substantial effect on fit indices. For example, as the N:q ratio increases, the CFI approaches 1.00 and the RMSEA approaches 0.00. However, he found that the absolute sample size has a much larger effect on fit indices than the N:q ratio. Besides the several rules of thumb that have been suggested, such as the ones presented above, some researchers have proposed methods for choosing a sample size based on the power it would provide to test a specific model (Kaplan, 1989, 1995; Lei, 2002; MacCallum, Browne, & Sugawara, 1996; Satorra, 1989; Satorra & Saris, 1985). For example, MacCallum et al. (1996) provides a method to determine sample size as a function of estimated effect size, degrees of freedom of the model, and desired power. He defines effect size in terms of the difference between the RMSEA value associated with the null hypothesis and any other specified RMSEA, and provides tables that associate sample size and power, given the effect size and degrees of freedom. Hancock and Freeman (2001) extended the tables provided by MacCallum et al. to several other combinations of values.

The issue of determining an adequate sample size for estimating latent growth models has not received much attention in the literature yet. Only a couple of very recent studies have tried to investigate how sample size affects the estimation of LGM. Hamilton et al. (2003) investigated the effect of sample size on convergence rates, chi-square statistic, fit indices, and parameter estimates. They found that the rate of convergence of a univariate LGM model with five measurement times is very high even with small sample sizes such as 50 and 100. With a sample size of 100, all conditions resulted in a convergence rate higher than 90%. Their results also showed that the relative bias of the chi-square values was consistently lower than 0.05 when the sample size reached 200. The researchers evaluated the effect of sample size on the comparative fit index (CFI) by looking at its mean across replications, which has an expected value of 1.00 because the model is correctly specified. CFI was higher than .99 with all conditions with sample size of 50. The effect of sample size on the standardized root mean squared residual (SRMR) and root mean squared error of approximation (RMSEA) were evaluated by comparing the means of these fit indices across replications with their expected value (0.00) for a correctly specified model. The average value of the SRMR decreased as sample size increased, and was below 0.05 with all conditions when sample sized reached 200. The RMSEA was found to decrease quicker than the SRMR and the mean RMSEA was below 0.05 in all conditions with the sample size of 50. The authors also investigated how sample size affects the relative bias of the parameters estimates. They found that the size of the sample does not bias the intercept, slope and covariance between intercept and slope. These estimates had mean relative bias across replications close to zero even with sample sizes as low as 50. The parameter estimate that presented the highest bias was the covariance between intercept and slope with sample size of 25. The variability of the parameter estimates reduced consistently as the sample size increased. Finally, sample size was found not to affect the standard errors of the parameter estimates. The authors concluded that for adequate estimation of correctly specified LGM models, a sample size of at least 100 is recommended. However, if there is some misspecification in the model, larger sample sizes may be needed.

Another study about sample size requirements in LGM was executed by Fan (2003a), who also compared LGM with repeated measures ANOVA with respect to power to detect group differences in intercept and slope. Fan modeled the group differences by including a dummy coded indicator in the LGM. The model studied was linear with five measurement times and eleven sample sizes (i.e. 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000). Fan simulated conditions where the groups were different with respect to the intercept only, slope only, and both intercept and slope. He also simulated conditions with interactions between group membership, intercept and slope. The standardized mean difference between groups in intercept and slope simulated were small (d = 0.2), medium (d = 0.5), large (d = 0.8) and no difference (d = 0.0).

Fan (2003a) found that the convergence rate of the LGM models was higher than 90% with sample sizes of 50 and higher than 99% with sample sizes of 100 or larger. With respect to type I error when there is no group difference, Fan found that both LGM and repeated measures ANOVA retain a type I error rate lower than 0.06 in all conditions. With respect to differences in slope, the author found that LGM shows more power than repeated measures ANOVA to detect group differences in slope, even with sample sizes as low as 50. However, the power of LGM to detect small group differences in slope only reaches 0.8 with sample sizes of 800. Sample sizes of 200 and 50 are needed for LGM to detect medium and large differences, respectively. LGM requires a sample size from one half to two thirds of the sample size required in repeated measures ANOVA to detect the same difference in the slope. Repeated measures ANOVA showed more power than LGM to detect differences in intercept if the differences are small. If the differences are medium and large and the sample size is larger than 300, there was no difference in power between the two methods. LGM was found to need sample sizes of 900, 200, and 50 to detect small, medium, and large differences of intercepts, respectively. When there was an interaction between the group membership and intercept and slope, the author found that LGM provides more power than repeated measures ANOVA to detect slope differences, with either small, medium or large differences. Repeated measures ANOVA could not be used to estimate differences in intercept when there was an interaction with group membership. With LGM, Fan found that sample sizes of 500 were needed to detect medium differences in slope when there was an interaction and a sample size of 700 was needed to detect differences in intercept.

No studies have been published investigating sample size requirements for the curve-of-factors model. However, the few applied studies found in the literature used moderate to large sample sizes. For example, Duncan and Duncan (1996) used a sample size of 321 individuals to fit a curve-of-factors model with a single factor measured at four time points with three indicators per factor and one predictor. Hancock, Kuo and Lawrence (2001) used a sample of 791 subjects to fit a model with three time points and three indicators per factor.

Statement of the problem

The literature review presented above has shown that modeling a composite of multiple indicators is not equivalent to modeling a latent variable, and ignoring the measurement error of the indicators in LGM of latent constructs may produce biased parameter estimates. Furthermore, it has been pointed out that univariate LGM assumes strict measurement invariance, and that the presence of other types of invariance or non-invariance may also introduce bias in the parameter estimates. In the previous sections, four methods a researcher can select to perform LGM of multiple indicators have been described: univariate LGM of composites of items, of composites whose variance/covariance matrix is corrected for attenuation, of composites with fixed error variances, and multivariate LGM with the curve-of-factors model.

The previous sections have also shown that there have not been many studies about sample size requirements for LGM in the literature. Furthermore, the SEM literature has provided mixed guidelines with respect to sample size and a consensual recommendation has not been reached yet.

In the context of analyzing the growth of latent variables measured by multiple indicators, this dissertation aims to evaluate the four LGM methods presented with respect to sample size requirements and the ability of these methods to provide unbiased estimates of the growth parameters (i.e. means and variances of the level and shape, and the covariance between level and shape) given the presence of measurement error and different invariance conditions. In order to accomplish this goal, samples of different sizes will be simulated, mimicking conditions that could be found with real data. Because the focus is on the effects of measurement error and measurement invariance in LGM, the population parameters that will be manipulated are the ones in the measurement model of the latent factors. These parameters are the factor loadings, measurement errors and intercepts of the indicators at each measurement time. However, an applied researcher who uses LGM is usually more interested in the parameters of the structural part of the model. Consequently, the main criteria to evaluate the performance of the four LGM methods studied in this dissertation will be whether these methods produce unbiased estimates of the means, variances and covariance of the level and shape factors and their corresponding standard errors. Furthermore, because information about the fit of hypothesized models is a very important part of applied research, the four methods will also be compared with respect to whether the fit indices resulting from the analyses with each method would lead to the retention or rejection of the model.

The research question that will be addressed in this dissertation with respect to measurement error is: How does the amount of measurement error in the indicators affect the results of the four methods of LGM? The research question about measurement invariance is: How does non-invariance and different types of measurement invariance affect the results the four methods of LGM? Finally, the research question about sample size is: Do the results from the four LGM methods differ with respect to sample size?

The questions about invariance are closely related to the ones about measurement error because strict factorial invariance assumes equal measurement errors across testing occasions, and strong, weak, and configural invariance relax this assumption. This means that a LGM method which fails to account for differences in measurement error across testing times may only produce unbiased parameter estimates if strict factorial invariance holds. This project does not address methods to test invariance, only the consequences of assuming a certain type of invariance when it does not exist. This information may be useful to inform researchers about which method is most appropriate for their longitudinal analysis of latent constructs.

It is hypothesized that univariate LGM of composites will only produce unbiased parameter estimates when there is strict factorial invariance and the reliability of the composites is high. It is also hypothesized that LGM of composites with the correction for attenuation, LGM of composites with fixed error variances, and the curve-of-factors model will provide unbiased parameter estimates in the conditions where the indicators have low reliability and there is strict or weak factorial invariance.

The information obtained in this dissertation about the effectiveness of different LGM methods to analyze longitudinal data from multiple indicators containing measurement error and varied degrees of invariance may guide applied researchers to choose the appropriate LGM method for their analysis. The section below details the conditions included in this study, the method to be used to simulate data, and the criteria to be used to evaluate the outcomes of the LGM methods.

Chapter III: Method

In this study, the effect of measurement error, different types of invariance, and sample sizes on the growth parameters of LGM of multiple indicators was evaluated. This was accomplished through a Monte Carlo simulation study where one thousand samples were created according to each condition. The conditions were designed to mimic situations encountered in applied studies, where researchers have to analyze data with different reliabilities, construct invariance, number of measurement occasions, number of indicators and sample sizes. In this section, first the conditions that were manipulated in this study are described, then the procedure chosen for data simulation and analysis is specified, and finally the method that was employed to evaluate the results is presented.

Methods of LGM analysis

The following alternatives for LGM of latent constructs measured by multiple items were evaluated:

1. LGM of item means: This is the method most commonly used in applied research. To implement this method, the means of all the items hypothesized to assess each latent construct were calculated for every individual, and a univariate LGM model was fit to the moment matrix of the means.
2. LGM of item means, with a correction for attenuation: In this method, the item means were calculated, and their correlation matrix was obtained. Cronbach's alpha reliability estimates were calculated for each group of items assessing the

latent construct at each measurement time. These reliability estimates were used to disattenuate the correlations (see equation 30) and variances (see equation 33) of the composites. The correlation matrix was transformed into a moment matrix by multiplying the dissatenuated correlation matrix by the square root of the dissatenuated variances of the composites and adding a vector of means. The resulting moment matrix was analyzed with univariate LGM.

3. LGM of item means, fixing the error variances: Reliability estimates $\rho_{C_jC_j}$ of the composites at each measurement occasion were calculated using the Cronbach's alpha formula. Next, the variances $\sigma^2_{C_j}$ of the composites at each measurement time were calculated. Then, the estimates of the error variances of the composites (i.e. $(1-\rho_{C_jC_j})\sigma^2_{C_j}$) were calculated. A univariate LGM model was fit to the moment matrix of composites, with error variances fixed at $(1-\rho_{C_jC_j})\sigma^2_{C_j}$.
4. The Curve-of-factors model: This model was fit to the moment matrix of the indicators. All factor loadings, error variances and intercepts of the indicators in the measurement model were freely estimated, with the exception of the loading and intercept of the first indicator at each measurement time, which were fixed to one and zero, respectively.

Population parameters

The five parameters of interest in latent growth modeling are the means and variances of the level and shape, and the covariance between level and shape. In this simulation, the means of the level and the shape were set to 1. The variance of the level was set to 0.5 and the variance of the shape was set to 0.1. Muthén and Muthén (2002) argue that a ratio between variance of the level and variance of the shape of 5 to 1 is commonly found in the applied literature. The correlation between the level and shape factors was set to 0.4.

A linear growth trajectory with equally spaced measurement occasions was specified for all conditions. To accomplish this, the loadings of the shape factor were fixed at increasing values starting from zero on the first measurement occasion, with increments of one. The loadings of the level factor were all fixed at one.

The parameters of the measurement part of the population model (i.e. factor loadings, error variances, and intercepts) were specified depending on the reliability, type of items and invariance conditions, which will be described in a subsequent section. In the population model, the variances of the latent factors at each measurement occasion (which in the curve-of-factors model are represented by the disturbances of the latent factors) were set to 0.2.

Sample size

The sample sizes simulated in this study were 100, 200, 500, and 1000. A sample size between 100 and 200 is the minimum recommended by Hamilton (2003) and Fan (2003a) for parameter estimation in univariate LGM. The ratio recommended by Kline (1998) of 10 observation for each parameter estimated (i.e. 10:1 ratio) was also used as a criterion to determine sample sizes. The smallest model (i.e. univariate LGM with three measurement times and with error variances constrained to $((1-\rho_{C_jC_j})\sigma^2_{C_j})$ has 8 parameters to be estimated, so the minimum required sample size would be 80. The largest model (i.e. curve-of-factors model with five measurement times and fifteen indicators per factor) has 90 parameters, so the minimum recommended sample size would be 900. Therefore, the sample sizes of 100 and 1000 were used to satisfy the criterion of 10 observations for each parameter to be estimated for the smallest and largest model, respectively. Fan (2003a) indicated that to detect small differences in means and variances of level and shape between two groups with a power of 0.8, a sample size of at least 500 is required, so a sample size of 500 was also simulated.

Number of items per factor

The number of items per factor affects the degrees of freedom, the power to detect a model misspecification, and the number of iterations necessary for convergence. Fan (Fan, 2003a) showed that with the same sample size, models with more items per factor converge more quickly. Furthermore, the reliability of the scores increases as the number of items increase. However, in this study the values of the reliabilities were fixed independently of the number of items, so that the effect of reliability and of the number of items could be analyzed separately.

In this study, latent factors with five, ten and fifteen indicators were used. This range of values aims to reflect three scenarios encountered in applied studies with respect to how a researcher might select indicators to measure a latent construct: In the first scenario, a researcher may use only some items to measure a single construct from a larger survey measuring several different constructs (e.g. the NELS is a large survey containing a few items measuring different latent constructs, such as motivation and parental involvement). In this scenario, the number of items selected is typically small (e.g. Fan, 2001). In the second scenario, a researcher chooses parts of an existing scale to measure the same latent construct measured by the entire scale (e.g. Mason, 2001). In the third scenario, the researcher uses an entire scale previously constructed or builds a scale specifically for the longitudinal study (e.g. Li, F. et al., 2001). Although these scenarios (more typically the second and third) can have a large number of items measuring a construct, this study included a maximum of fifteen items due to the difficulty of manipulating large variance/covariance matrices of items.

Type of items

Two types of items were simulated: essentially tau-equivalent and essentially congeneric. The objective of simulating these two types of items relates to the fact that Cronbach's alpha reliability estimates are used in two of the methods reviewed in this study (the correction for attenuation and LGM method with fixed error variances). Cronbach's alpha assumes that the items are essentially tau-equivalent (Lord & Novick, 1968) and has been shown to underestimate reliability if the items are congeneric (Komaroff, 1997; Osburn, 2000; Raykov, 1997, 1998, 2001).

In order to thoroughly investigate the performance of the two methods that are based on Cronbach's alpha, datasets were simulated with either essentially tau-equivalent or essentially congeneric items. The difference between tau-equivalent and essentially tau-equivalent is that in the latter the items have different intercepts. Although the term essentially tau-equivalent has been in use for some decades (Lord & Novick, 1968), the term essentially congeneric was proposed more recently by Millsap and Everson (1991) to describe a condition where items have different factor loadings, error variances, and intercepts. This study simulated essentially congeneric items instead of congeneric ones because it is a more general condition.

The classification of items into parallel, essentially parallel, tau-equivalent, essentially tau-equivalent, congeneric, and essentially congeneric (Millsap & Everson, 1991) refers to the similarity of the characteristics (i.e. factor loadings, error variances, and intercepts) of different items that measure the same construct at a single testing occasion. This classification is only meaningful if the construct is measured by at least two items, which can be compared to each other within a testing occasion. On the other hand, measurement invariance refers to the stability of the characteristics of each individual item measuring a construct at a single testing occasion as compared with the same item in another testing occasion. Consequently, it is possible to talk about the invariance of a single item.

In the essentially tau-equivalent condition, all the population loadings of the indicators in every measurement time were set to one, but the error variances were set to different values. The population values of the loadings, intercepts, and error variances for all conditions are presented at the appendix. The existence of tau-equivalent items also has implications with respect to invariance. If the items are tau-equivalent in every testing occasion, the condition of weak factorial invariance is also met. Consequently, it is not possible to have items that are both essentially tau-equivalent and non-invariant.

In the essentially congeneric condition, the items had different loadings on the latent factor. Because the effect of a specific choice of factor loadings is not of interest in this study, different values of population factor loadings for each indicator were randomly chosen from a uniform distribution within a range of values that could be found in applied research. In this simulation, factor loadings were drawn within the range of 0.5 to 1, rounded to two decimal places. The mean factor loading was 0.75 and the standard deviation was 0.14. This randomization of factor loadings averaged any effect that the choice of factor loadings could have had in the results. In order to guarantee that the essentially congeneric condition was created, the factor loadings randomly drawn were examined to verify that they differed from each other. Whether the loadings randomly chosen for the set of indicators of one latent variable varied across measurement times depended on the invariance conditions simulated. The process of fixing or varying the loadings to reflect different invariance conditions will be described in a later section. The population loadings for all conditions included in this study are presented in appendix C.

The population intercepts of the indicators of each factor in both the essentially tau-equivalent and essentially congeneric conditions were defined by randomly sampling a number from a uniform distribution with a minimum of 1, maximum of 10, mean of 5.5, and standard deviation of 2.6. The reason to choose intercepts randomly in a limited range of numbers is to keep them in a scale similar to the population means of the level and shape factors, but at the same time create random variability of the intercepts across conditions. Whether the values chosen for the intercepts at a testing occasion varied across time depended on the invariance condition simulated. The population intercepts for all conditions are reported in appendix C.

Both essentially tau-equivalent and essentially congeneric items have unequal error variances. The error variances of each indicator were also randomly chosen, but with the constraints that given the number of items and the factor loadings chosen, the measurement errors produced the desired composite reliability values. More details about the specification of measurement errors will be given in the section about reliability.

Reliability

One of the focuses of this study is on the effects of measurement error. Measurement errors of the indicators forming a composite were specified through the composites' reliabilities because reliability coefficients are commonly reported in the literature, and could be used to provide realistic reference values for the simulation. Many reliability generalization studies have been recently published in the literature, and these studies provide the mean reliability coefficient across several studies about a specific latent construct. For example, reliability generalization studies have reported that the mean Cronbach's alpha of scales to measure social desirability bias (Beretvas, Meyers, & Leite, 2002), geriatric depression (Kieffer & Reese, 2002), state anxiety and trait anxiety (Barnes, Harp, & Jung, 2002), and life satisfaction (Wallace & Wheeler, 2002) were .72, .84, .91, .89, and .79, respectively. In this study, the reliability of the composites simulated was set to 0.7 and 0.9, which correspond to low and high values of mean reliability coefficients found in the reliability generalization studies reviewed.

The reliability $\rho_{C_j C_j'}$ of a composite is a function of the error variances, the factor loadings and the number of indicators of each factor, according to the formula below, which was presented previously in the discussion about measurement error:

$$\rho_{C_j C_j^{'}} = \frac{\left(\sum_{k=1}^{K} \lambda_k\right)^2}{\left(\sum_{k=1}^{K} \lambda_k\right)^2 + \sum_{k=1}^{K} \sigma_{\varepsilon_k}^2} \tag{49}$$

where λ_k is the factor loading of indicator *k* on a latent factor with *K* indicators, and $\sigma_{\varepsilon_k}^2$ is the error variance of indicator *k*.

In order to obtain the indicators' error variances that correspond to each composite's reliability, the formula above was used to solve for the sum of the error variances $\sum_{k=1}^{K} \sigma_{\varepsilon_k}^2$ as a function of each reliability value and the randomly chosen factor loadings. Because both essentially tau-equivalent and essentially congeneric items have different error variances, the sum of error variances was divided unequally among the items of each latent factor. This division was accomplished by randomly selecting *k* real numbers in the 0 to 10 range, where *k* is the number of indicators of each factor, then dividing each number selected by the sum of *k* numbers, and multiplying the fractions obtained by the sum of error variances. This method gave to each indicator a randomly chosen part of the sum of error variances, and guaranteed that the composite reliability was equal to the value specified.

Whether the error variances calculated for the indicators of a latent factor at one testing occasion varied on the other testing occasions depended on the invariance condition simulated. Details on how measurement errors, factor loadings and intercepts are specified to reflect each invariance condition are given in the next section.

Invariance

Three invariance conditions were simulated: strict factorial invariance, weak factorial invariance, and configural invariance.

In the strict factorial invariance condition, the population factor loadings, error variances, and intercepts were set to be the same across testing times. This was accomplished by first choosing the factor loadings, error variances, and intercepts for the first measurement time using the methods described in the previous sections, according to the type of item (i.e. essentially tau-equivalent or essentially congeneric) and construct reliability. Then, strict factorial invariance was obtained by specifying the same population item parameters of the first measurement time for the other measurement times.

In the weak factorial invariance condition, only the population factor loadings were identical across testing times. The population factor loadings chosen for the first measurement time were set to the same values in the other measurement times. The population error variances and intercepts of each item were set to different values across conditions.

Configural invariance was modeled by setting the population factor loadings, error variances and intercepts to different values across testing times, but the signs of the factor loadings remained the same. Configural invariance was simulated because it is a more realistic type of non-invariance than the condition where the signs of the loadings changes across testing times or some of the loadings become zero.

Number of measurement times

The number of waves of data collection has a important role in LGM because the precision of parameter estimates tends to increase along with the number of observations for each individual (Duncan, T. E. et al., 1999). Also, more measurement times allow more flexibility in modeling the growth shape. The collection of data in just two waves allows only for linear change, while if multiple waves of data are collected, other growth shapes can be estimated. In this study, data was simulated with three and five measurement occasions.

Study design overview

Data for this study were simulated with four sample sizes (100, 200, 500, 1000), three invariance conditions (strict, strong and configural invariance), two reliability values (0.7 and 0.9), two levels of number of measurement occasions (3 and 5), and with 5, 10 or 15 items per factor which were either essentially tau-equivalent and essentially congeneric. The conditions were not completely crossed because the condition with essentially tau-equivalent items implies at least weak factorial invariance. Consequently, there were no conditions with essentially tau-equivalent items and configural invariance. This simulation study had 144 conditions with essentially congeneric items and 94 conditions with essentially tau-equivalent items, which sums to 240 conditions.

In addition to the conditions mentioned above, some "ideal" conditions were created where the reliability was perfect (i.e. measurement error was zero), the items were parallel (i.e. equal loadings, intercepts and zero error variances) and there was strict factorial invariance. There were only 24 of these conditions, reflecting combinations of number of items (i.e. 5, 10, and 15), number of measurement times (i.e. 3 and 5) and sample size (i.e. 100, 200, 500, and 1000). These conditions were created to serve as a baseline to compare the LGM methods, because all of them should perform well under these "ideal" conditions.

Four methods (LGM of item means, LGM of item means with the correction for attenuation, LGM of item means with fixed error variances, and the curve-of-factors model) were used to analyze datasets generated under each of the conditions. A summary of the conditions included in this study can be found in Appendix A. A summary of the ideal conditions can be found in Appendix B. The population values of all the parameters used in the simulation can be found in Appendix C.

Data generation

First, the matrix equations of the curve-of-factors model were filled with the population values of the growth parameters (i.e. means and variances of level and shape, covariance between level and shape, and disturbances of latent factors) and the measurement parameters (factor loadings, measurement errors, and intercepts). The matrix equations of the curve-of-factors model were used to generate data because this model contains all the parameters, and the univariate LGM models are simplifications of it. For each condition, the matrix equations of the curve-of-factors model with the parameters filled in were solved with the software R 2.0.1 (R Development Core Team, 2004) to obtain the population variance/covariance matrix of the items and the vector of item means.

In the next step, the population variance/covariance matrix of the items and the vector of item means for each condition were used in a *R 2.0.1* function to generate multivariate normally distributed random numbers representing the scores of all individuals on each item. Each dataset contained scores that were random and normally distributed deviates of the population values specified in the variance/covariance matrix and vector of means of the items. One thousand datasets was generated for each of the 264 conditions, with a different random seed for each condition.

Once the 264,000 datasets were created, they were saved to the disk, and *R 2.0.1* was used to perform the first steps of the latent growth analyzes. Because two of the LGM methods that were evaluated in this study use reliability estimates, the first step was to calculate Cronbach's alpha reliability estimates of the set of items within each testing time.

Because the three univariate LGM methods evaluated in this study are based on the analysis of composites (i.e. means) instead of individual item scores, *R 2.0.1* was used to calculate the means of the items for each subject within each testing time, and the composites' variance/covariance matrix and vector of means. For the univariate LGM method with the correction for attenuation, *R 2.0.1* was used to calculate correlation matrices instead of variance/covariance matrices, which were disattenuated using the estimated Cronbach's alpha coefficients for each composite. After disattenuation, these correlation matrices were multiplied by the square root of the disattenuated variances of the composites to obtain the composites' variance/covariance matrices.

In order to implement the curve-of-factors model, *R 2.0.1* was used to calculate the variance/covariance matrix of the items, and their vector of means for each dataset.

After all the input matrices for the four LGM methods were created, these matrices were saved to the disk and the analysis proceeded to the model fitting stage.

The software Mplus 3.11 and the Runall utility (Muthén & Muthén, 2004) was used to fit the univariate LGM models and the curve-of-factors model to every variance/covariance matrix and vector of means. The Runall utility allows Mplus to repeatedly fit SEM models to data that were generated outside of Mplus.

In every model, the factor loadings of the level were fixed to one and the factor loadings of the shape were fixed to values increasing by one unit starting at zero. Also, the growth parameters (means and variances of the level and shape, and their covariance) were freely estimated in all the models. In the LGM of item means and LGM of item means with the correction for attenuation the error variances of the composites were freely estimated. In the LGM of item means with the error variances fixed at $(1-\rho_{C_jC_j^{'}})\sigma_{C_j}^2$, the disturbances of the latent factors at each measurement time were freely estimated. This disturbance corresponds to the reliable variance of the composite. The reliability coefficient, $\rho_{C_jC_j^{'}}$, was the sample Cronbach's alpha calculated with *R 2.0.1* for each composite at each measurement time. In the curve-of-factor models, the factor loadings, error variances, and intercepts of each item were freely estimated, in addition to the growth parameters. However, the first loading of each factor was constrained to one in order to set the scale of the factors.

The four LGM models were fit to the 1,000 datasets of each of the 264 conditions, and the convergence rates and percentages of inadmissible solutions were recorded. The initial set of datasets was analyzed without removing the inadmissible solutions. Next, inadmissible solutions were removed and additional datasets were simulated until 1,000 admissible solutions were obtained for each condition with each method. The reason for obtaining 1,000 sets of results for each condition with each method is that the LGM methods were compared with respect to the means and standard deviations of the parameter estimates and fit indices obtained across replications. In order to fairly compare the LGM methods, the means and standard deviations of the parameter estimates and fit indices should be obtained from the same number of observations (i.e. replications).

After fitting the different LGM models, Mplus outputs the parameter estimates, standard errors of parameter estimates, and fit statistics. These results were saved to the disk so they could be used in the data analysis stage to compare the LGM methods.

Data analysis

The first step in comparing the performance of the four LGM methods was to look at their convergence rates. The convergence rate of each LGM method was evaluated by comparing the number of analyses (from a total of 1000) that converged to a solution. The average convergence rate across replications for each method with each condition was reported.

The next step was to compare the four LGM methods with respect to the bias in the parameter estimates. Bias in the following five parameter estimates was considered:

mean of the level, mean of the shape, variance of the level, variance of the shape, and covariance between level and shape. The comparison of the LGM methods was performed using estimates of relative bias (Hoogland & Boomsma, 1998) averaged across replications. This evaluation of the LGM methods is based on the concept that if a method is unbiased, the expected value of the difference between the estimate and the population parameter is zero. The relative bias is the difference between the parameter estimate $\hat{\theta}_{rc}$ for each replication r in each condition c and the generating population parameter θ, divided by the population parameter. In other words, the relative bias is the ratio of the bias and the population parameter. The formula for the relative bias is:

$$B(\hat{\theta}_{rc}) = \frac{\hat{\theta}_{rc} - \theta}{\theta} \tag{50}$$

The relative bias is considered acceptable if its magnitude is less than .05 (Hoogland & Boomsma, 1998). The one thousand relative bias estimates for each condition were used in a quantitative analysis investigating the effect of the conditions on the magnitude of the relative bias. This quantitative analysis was performed through several ANOVAs, one for each of the five parameter estimates and for each of the four methods of latent growth modeling. In these six-way ANOVAs, the dependent variable was the relative bias of a parameter estimate for all conditions with each LGM method, and the independent variables were invariance condition, type of item, sample size, reliability, number of items per factor, and number of measurement times.

The effect sizes of the conditions included in the ANOVAs was examined using the partial Eta-squared. The partial Eta-squared is an index of effect size obtained by dividing the sum of squares of an effect by the sum of squares error plus the sum of squares of the effect (i.e. $SS_{effect} / (SS_{error} + SS_{effect})$). The partial Eta-squared differs from the Eta-squared because the latter is the ratio of the sum of squares of an effect and the sum of squares total. With the Eta-squared, if the same experiment is repeated with some extra predictors, the value of Eta-squared of each predictor will be different from the previous experiment, because the sum of squares total will change. The partial Eta-squared has the advantage over the Eta-squared of staying constant for one predictor across two different studies. The partial Eta-squared is similar to the partial correlation in the sense that it is a index where for each predictor, the effects of the other variables are removed (Cohen, 1973). The Eta-squared does have one advantage over the partial Eta-squared: Within an experiment, the Eta-squared of the predictors sum to one. However, as Cohen (1973) emphasizes, it is a trade-off for the property of the partial Eta-squared of being useful to compare effect sizes across studies. Because this study is based on simulations, it is desirable to report an effect size index which can be used to make comparisons with future replications and expansions of the study. Only partial Eta-squared coefficients equal or larger than 0.05 were reported.

Because the standard errors of the parameter estimates are essential to test hypotheses about the parameter estimates, and to calculate their confidence intervals, the four methods of LGM were also compared with respect to relative bias of standard error estimates. The relative bias of the standard errors was evaluated for means and variances of the level and shape, and the covariance between level and shape. The relative bias of standard errors can be calculated with the formula (Hoogland & Boomsma, 1998):

$$B\left(\hat{s}e_{\theta_{rc}}\right) = \frac{\hat{s}e_{\theta_{rc}} - \hat{s}_{\theta_c}}{\hat{s}_{\theta_c}} \tag{51}$$

where $\hat{s}e_{\theta rc}$ is the estimated standard error of a parameter in replication *r* of condition *c*, and $\hat{s}_{\theta c}$ is the standard deviation of the one thousand parameter estimates of condition *c*, which is an estimate of the population standard error. The bias of the standard error estimates is considered acceptable if its magnitude is less than 0.1 (Hoogland & Boomsma, 1998). The same six-way ANOVAs that were conducted with the bias of the parameter estimates were also performed with the bias of the standard errors.

In structural equation modeling, a fundamental issue that needs to be investigated before interpreting any estimated parameter is whether the model fits the data. In the SEM framework, model fit means whether the implied variance/covariance matrix is close to the observed variance/covariance matrix (Bollen, 1989). Assessment of model fit is also an important part of latent growth modeling. In this project, the four LGM methods were evaluated with respect to whether they produced unbiased fit statistics (i.e. Chi-square fit statistic) and fit indices that indicated an acceptable fit of the model. Overall model fit indices were used, because they evaluate the model as a whole. The three overall model fit indexes used in this study were the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean-square error of approximation (RMSEA), because they are commonly used to determine the acceptability of models in applied studies, and criteria about which values of the CFI, TLI, and RMSEA correspond to well-fitting models have been well established (see Hu & Bentler, 1999).

The adequacy of the chi-square statistics provided by each LGM method was evaluated by calculating its relative bias. The relative bias of the chi-square is equal to the difference between the chi-square statistic, χ^2_{rc}, of each replication *r* of condition *c* and the expected chi-square $E[\chi^2_c]$ of condition *c*, divided by the expected chi-square. For a correctly specified model, the expected chi-square is equal to the degrees of freedom of the model.

$$B(\chi^2_{rc}) = \frac{\chi^2_{rc} - E[\chi^2_c]}{E[\chi^2_c]} \quad (52)$$

The relative bias of the chi-square has been considered acceptable if its magnitude is less than 0.05 (Hamilton et al., 2003). This study evaluated the performance of the methods with respect to the chi-square statistic by computing its relative bias instead of computing the percentage of replications where the model would be retained based on the chi-square statistics, because the chi-square statistic is a very strict criterion of fit (i.e. it indicates exact fit), and is very sensitive to small degrees of misfit. Furthermore, six-way ANOVAs were conducted to examine the effect of the conditions included in this study on the relative bias of the chi-square statistic.

The adequacy of the CFI, TLI, and RMSEA produced by the four LGM methods was evaluated by looking at the percentage of replications whose CFI, TLI and RMSEA would indicate acceptable fit. Values of CFI and TLI equal or higher than 0.95 and values of RMSEA equal or lower than 0.05, are considered to indicate acceptable fit (Hu & Bentler, 1999).

The relationship between the conditions manipulated in this study and the CFI, TLI and RMSEA was quantified by calculating the Cramer's V coefficient for each pair of variables formed by a condition (i.e. number of items, number of measurement times, sample size, reliability, invariance, and type of item) and a dummy coded variable indicating whether or not the CFI, TLI and RMSEA support acceptable model fit. For contingency tables, the chi-square statistic allows a test of the null hypothesis that the frequencies of two variables are independent, but does not provide information about the strength of association between the variables (Acock & Stavig, 1979; Agresti, 1996). The Cramer's V, however, provides a measure of the strength of the relationship between two categorical variables (Acock & Stavig, 1979; Argyrous, 2000; Hays, 1973). The Cramer's V is calculated based on the chi-squared statistic:

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{\chi^2_{\mathbf{max}}}} \tag{53}$$

Where χ^2 is the chi-square statistic obtained from a contingency table and $\chi^2_{\mathbf{max}}$ is the maximum value of the chi-square statistic when the variables are completely dependent. The value of $\chi^2_{\mathbf{max}}$ is $n[MIN(r-1,c-1)]$, where $MIN(r-1,c-1)$ refers to the smaller of either the number of rows *r* or the number of columns *c*. Cramer's V has the advantage of being constrained between zero and one and having a clear proportional interpretation (Acock & Stavig, 1979):

$$\text{Cramer's V} = \frac{\text{Obtained departure from independence}}{\text{Maximum departure from independence}}$$

Because the number of replications simulated in this study was very large, even small relationships between variables would be statistically significant. Therefore, only conditions that produced a Cramer's V equal or larger than 0.1 were reported.

Chapter IV: Results

This chapter presents the results of the simulation study performed to compare four methods of growth modeling of latent constructs measured by multiple indicators. The results chapter is divided into six sections. In the first four sections, the results of the curve-of-factors model, univariate LGM, univariate LGM with fixed error variances, and univariate LGM with correction for attenuation are presented, in this order. The fifth section compares the results of the four methods. The sixth section provides the results of the four methods with conditions where reliability was perfect (i.e. measurement error was zero), there was strict factorial invariance and the items were parallel. There are twenty four of these "ideal" conditions, because the only factors manipulated were sample size (i.e. 100, 200, 500, and 1000), number of items (i.e. 5, 10, and 15), and number of measurement times (i.e. 3 and 5), while reliability, invariance and type of item were kept constant. The results for these ideal conditions are reported separately from the other results because they would make the design not completely crossed.

The four methods were compared with respect to convergence rates, relative bias of the estimates of the means, variances and covariance of level and shape, and their standard errors, relative bias of the chi-square statistic, and percentage of fit statistics (i.e. CFI, TLI and RMSEA) that would lead to retention of the model.

The results of the ANOVAs of the relative biases are presented for conditions for which the ANOVAs indicated effect sizes larger than 0.05 (as assessed by partial Eta-squared). For each method, this chapter indicates whether the relative biases were acceptable under the combination of conditions, and whether the relative biases were consistent across conditions or a particular combination of conditions presented larger relative biases than the others. Tables of relative bias are displayed according to combinations of conditions that showed differences in the acceptability of the relative bias, collapsing across the conditions which did not show any difference in the acceptability of relative bias.

The curve-of-factors model

The curve-of-factors model presented few problems of convergence, even with sample sizes as low as one hundred. From the analyses of the 240,000 datasets simulated (240 conditions with 1,000 datasets each) where the curve-of-factors model was fit to the data, only 27 analyses did not converge, which corresponds to a 99.99% convergence rate.

Although non-convergence was not a problem, many conditions resulted in estimated variance/covariance matrices of the level and shape that were non-positive definite. The non-positive definite matrices were due to either a negative estimate of the variance of level or shape, or an estimated correlation between level and shape outside of the -1 to 1 range. Because the maximum likelihood estimator used in structural equation modeling programs produces unbounded estimates (Wothke, 1993), the estimated variances and covariances may be inadmissible. The overall percentage of inadmissible solutions was 21.3% with three measurement times, but only 0.3% with five measurement times. Furthermore, the percentage of inadmissible solutions decreased as the sample size increased, as shown in Table 1. More details about the reasons for these inadmissible solutions are presented in a later section.

Table 1
Percentage of inadmissible solutions with the curve-of-factors model

Sample size	Times		Total
	3	5	
100	32.4%	0.9%	16.7%
200	25.2%	0.1%	12.7%
500	14.9%	0%	7.5%
1000	8.9%	0%	4.4%

Because removing the inadmissible solutions changes the distribution of the parameter estimates, in order to provide a complete picture of the ability of each method to produce unbiased parameter estimates, the relative bias of all parameter estimates was calculated twice: The relative bias was calculated once without removing the inadmissible solutions. Then, the relative bias was computed a second time after removing the inadmissible solutions.

In this study, the removal of inadmissible solutions affected more strongly the distributions of the variances of level and shape. However, the relative bias with and without removing admissible solutions will be reported for all parameter estimates, to allow their comparison. After the inadmissible solutions were removed, extra datasets were simulated to obtain one thousand admissible solutions for each condition.

When the relative biases of the variances of the level and shape are computed without removing the inadmissible solutions, the distribution of the variances has two tails, as shown in Figure 8. However, when the inadmissible solutions are removed, the distribution of the variances loses its left tail, as shown in Figure 9, because all estimates of variance smaller than zero are eliminated.

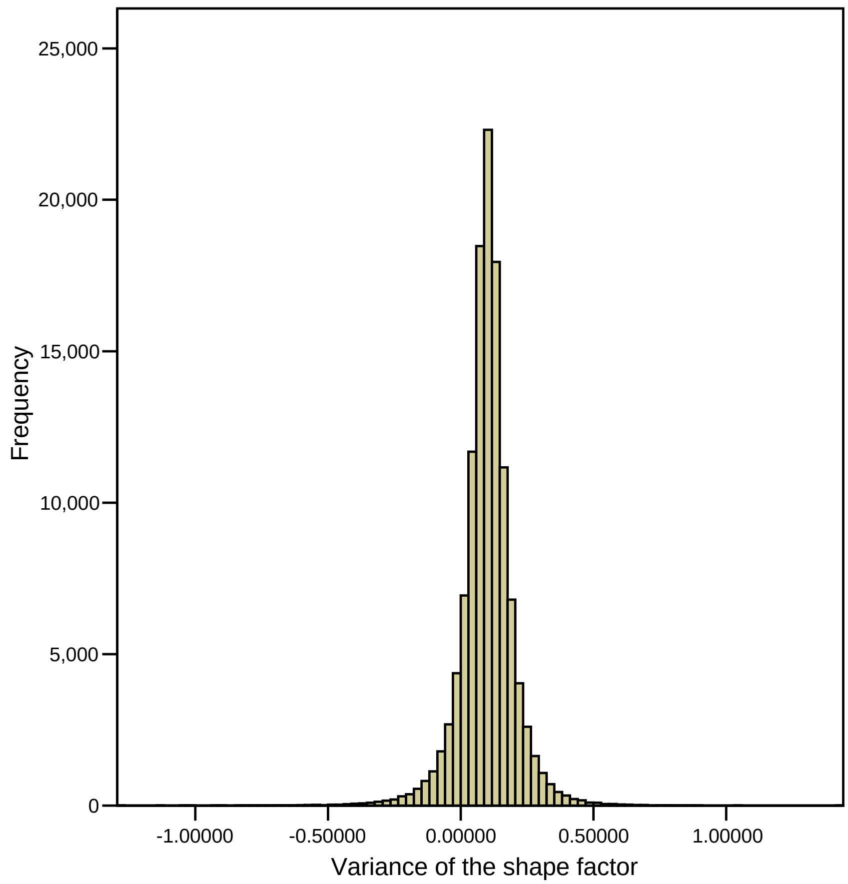

Figure 8. Distribution of the variance of the shape factor with three measurement times, including inadmissible solutions.

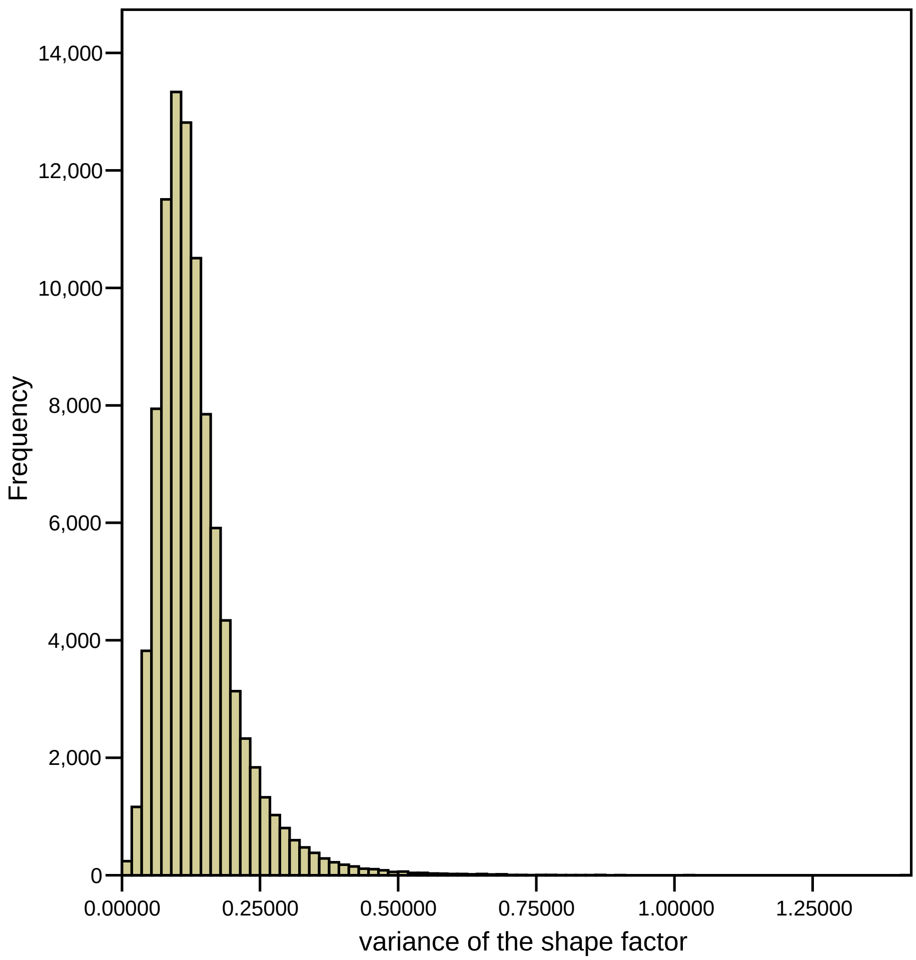

Figure 9. Distribution of the variance of the shape factor with three measurement times, excluding inadmissible solutions.

The definition of an unbiased estimator is one where the expected value of the distribution of estimates is equal to the parameter estimated (i.e. $E(\hat{\theta}) = \theta$) (Wackerly et al., 1996). This definition requires that, in calculating the relative bias of the parameter estimates, the entire distribution of parameters is considered (i.e. the distribution containing both tails).

For the curve-of-factors model, when the relative biases were calculated without removing the inadmissible solutions, the relative biases of the estimates of the means and variances of the level and shape, as well as the relative biases of their estimated standard errors were acceptable under all conditions. The ANOVA results indicated that none of the conditions had an effect on the relative bias of the estimates and on the relative bias of the estimated standard errors with a partial Eta-squared greater than 0.05. The relative bias of the estimate of correlation between level and shape cannot be calculated without removing the inadmissible solutions, because it is not possible to transform the estimated covariances into correlations when the variance estimates are negative. The relative bias of the estimate of the covariance between level and shape was reported instead of the relative bias of the estimate of the correlation. The relative bias of the covariance estimate was acceptable under all conditions with the exception of the condition when the number of measurement times was five, the number of items was fifteen and the sample size was one hundred. None of the conditions were associated with a partial Eta-squared value greater than 0.05 for either the estimate of the covariance or the estimate of the standard error of the covariance between level and shape. The relative biases of parameter estimates, without removing inadmissible solutions, are presented in table 2. The relative biases of standard errors, without removing inadmissible solutions, are presented in table 3.

Table 2
Mean relative bias of parameter estimates with the curve-of-factors model, without removing the inadmissible solutions, collapsing across reliability, type of item, and invariance conditions.

Times	Items	Sample size	Mean level	Mean shape	Variance level	Variance shape	Cov. level shape
3	5	100	.003	-.001	-.013	-.028	.006
		200	.000	.000	-.006	-.016	.004
		500	-.001	.000	-.006	-.013	.007
		1000	.000	.000	.000	.000	-.005
	10	100	.000	.000	-.012	-.006	.002
		200	.001	.000	.001	.011	-.019
		500	.000	.000	-.002	-.008	.004
		1000	.000	.000	.001	.008	-.007
	15	100	.002	.000	-.011	-.017	-.007
		200	.001	.000	-.006	-.004	-.010
		500	-.002	.001	-.005	-.004	.006
		1000	.000	.000	-.002	.004	.000
5	5	100	.001	-.001	.002	.000	-.027
		200	-.001	.001	-.001	.003	-.014
		500	.000	.000	-.002	.000	.000
		1000	.000	.000	-.001	.000	-.001
	10	100	-.001	.000	.010	.004	-.039
		200	.000	.000	.003	.004	-.023
		500	-.001	.000	.003	.001	-.012
		1000	-.001	.000	.002	.002	-.005
	15	100	.000	-.001	.009	.014	-.061
		200	-.001	.001	.004	.013	-.031
		500	.001	-.001	.002	.004	-.008
		1000	-.001	.000	.003	.001	-.006

Note. Cov. = covariance

Table 3
Mean relative bias of the estimates of the standard errors with the curve-of-factors model, without removing the inadmissible solutions, collapsing across reliability, type of item, and invariance conditions.

Times	Items	Sample size	Mean level	Mean shape	Variance level	Variance shape	Cov. level shape
3	5	100	-.010	-.010	-.026	-.026	-.020
		200	-.004	-.005	-.011	-.010	-.013
		500	.000	.005	-.005	.001	-.001
		1000	-.004	.007	.002	-.001	-.013
	10	100	-.019	-.005	-.014	-.028	-.018
		200	.005	-.001	-.021	-.011	-.008
		500	.004	-.009	-.001	-.006	-.007
		1000	.004	.009	-.012	-.012	-.008
	15	100	.001	-.013	-.026	-.018	-.019
		200	-.018	-.006	-.011	-.005	-.014
		500	-.008	.007	.000	.010	-.002
		1000	.001	.007	-.009	-.022	-.002
5	5	100	-.018	-.018	-.022	-.024	-.033
		200	.004	-.005	-.009	-.020	-.017
		500	.004	-.004	-.013	-.013	-.004
		1000	-.003	.011	.008	-.003	-.002
	10	100	-.009	-.007	-.023	-.031	-.038
		200	-.012	-.007	-.012	-.003	-.017
		500	-.011	-.006	-.005	-.008	.001
		1000	-.006	-.008	-.005	.000	-.007
	15	100	-.016	-.002	-.028	-.036	-.030
		200	-.011	-.016	-.007	-.013	-.007
		500	.014	.000	-.002	-.010	-.013
		1000	-.007	-.003	-.018	.010	-.003

Note. Cov. = covariance

When the inadmissible solutions were removed, the curve-of-factors model produced unbiased estimates of the mean of the level and of the mean of the shape under all conditions (see Table 4). In the ANOVA results, none of the Eta-squared values were greater than 0.05. Under selected conditions, however, there was bias in the estimates of the remaining three parameters of interest: variance of the level, variance of the shape, and correlation of the level and shape. There was no bias in these three estimates under the condition of five measurement times, but there was substantial bias in these estimates when data included only three measurement times. Although under selected combinations of conditions the relative bias was unacceptable, none of the conditions had an effect on the relative bias of the variance of the level associated with a partial Eta-squared greater than 0.05. For the variance of the level, bias was largest (the parameter estimate was 21% too large) when there were only three measurement times, a small sample size (i.e. 100) and many items (i.e. 15). There was little substantial bias when sample sizes were 500 or 1,000 regardless of the number of items. For the estimates of the variance of the shape, bias depended on measurement time (partial η^2 = 0.1), sample size (partial η^2 = 0.55), and the interactions of those effects (partial η^2 = 0.61). The estimates of the variance of the shape were positively biased by as much as 92% when there were only three measurement times, a small sample size (i.e. 100) and many items (i.e. 15). Under the three measurement time condition, the only situation that resulted in a lack of bias in the estimates of the variance of the shape was with a sample size of 1,000 and only five items per construct. Finally, for the estimate of the correlation of the shape and level, bias depended only on the number of measurement times (partial η^2 = 0.57). The estimate of the correlation of the level and shape was negatively biased by as much as 72% when there were only three measurement times, a small sample size (100) and many items (15). Under the three measurement time condition, the only situations that resulted in a lack of bias in the estimate of the correlation of the level and shape was with a sample size of 1,000 or with a sample size of 500 and fewer items (5 or 10).

Table 4
Mean relative bias of parameter estimates across conditions with the curve-of-factors model, with the inadmissible solutions removed, collapsing across reliability, type of item, and invariance conditions

Times	Items	Sample size	Mean level	Mean Shape	Variance level	Variance shape	Cor. level shape
3	5	100	0.002	-0.001	0.115	0.510	-0.440
		200	0	0	0.079	0.331	-0.270
		500	-0.001	0	0.021	0.102	-0.036
		1000	0	0	0.008	0.036	0.022
	10	100	-0.001	0	0.183	0.794	-0.615
		200	0.001	0	0.099	0.418	-0.357
		500	0	0	0.032	0.144	-0.084
		1000	0	0	0.023	0.108	-0.040
	15	100	0.001	-0.001	0.214	0.915	-0.717
		200	0.001	0	0.118	0.507	-0.426
		500	-0.002	0.002	0.064	0.270	-0.191
		1000	0	0	0.031	0.133	-0.053
5	5	100	0.001	-0.001	0.004	0.002	0.021
		200	-0.001	0	-0.001	0.003	0.013
		500	0	0	-0.001	-0.001	0.011
		1000	0	0	-0.001	0	0.005
	10	100	-0.001	0	0.012	0.006	0.008
		200	0	0	0.002	0.004	0.007
		500	0	0	0.002	0	0
		1000	-0.001	0	0.002	0.002	0
	15	100	-0.001	0	0.011	0.016	-0.008
		200	-0.001	0.001	0.004	0.011	0.002
		500	0.001	-0.001	0.001	0.003	0.004
		1000	-0.001	0	0.003	0.001	0

Note. Cor. = correlation

When the inadmissible solutions were removed, the relative biases of the estimates of the standard errors obtained with the curve-of-factors model followed the same pattern as the relative biases of the parameter estimates, with the exception that, with three measurement times, the estimates of the standard errors of the variance of the shape and of the correlation between level and shape were only acceptable with 1,000 observations and five items (see Table 5). Given that relative biases of standard errors are considered acceptable if they are equal to or smaller than 0.1 (Hoogland & Boomsma, 1998), the curve-of-factors model produced unbiased estimates of the standard errors of the mean of the level and of the mean of the shape under all conditions (see Table 5). None of the conditions had an effect on the estimates of the standard errors of the means of the level and shape with an effect size higher than 0.05. Under selected conditions, however, there was bias in the estimates of the standard errors of the remaining three parameters of interest: variance of the level, variance of the shape, and correlation of the level and shape. There was no bias in these three standard error estimates under the condition of five measurement times, but there was some positive bias in these estimates when data included only three measurement times. For the estimate of the standard error of the variance of the level, the bias depended only on the number of measurement times (partial η^2 = 0.087). Bias was around 11% when there were only three measurement times and small to moderate sample sizes (i.e. 100, 200, 500). With sample sizes of 1000, the bias of the estimate mean of the level was acceptable. This pattern of bias was consistent across number of items. For the estimate of the standard error of the variance of the shape, the bias depended on measurement time (partial η^2 = 0.386) and the interaction between number of measurement times and sample size (partial η^2 = 0.087). The estimate of the standard error of the variance of the shape was positively biased by as much as 38% when there were only three measurement times, a small sample size (i.e.100) and many items (i.e.15). Under the three measurement time condition, the only situation that resulted in a lack of bias in the estimate of the standard error of the variance of the shape was with a sample size of 1,000 and only five items per construct. Finally, for the estimate of the standard error of the covariance of the shape and level, the bias depended on number of measurement times (partial η^2 = 0.318) and the interaction between number of measurement times and sample size (partial η^2 = 0.057). The estimate of the standard error of the covariance of the level and shape was positively biased by as much as 32% when there were only three measurement times, a small sample size (i.e.100) and many items (i.e.15). Under the condition with three measurement times, the only situation that resulted in a lack of bias in the estimate of the correlation of the level and shape was with a sample size of 1,000 and only five items.

Table 5
Mean relative bias of the standard errors across conditions with the curve-of-factors model, with the inadmissible solutions removed, collapsing across reliability, type of item, and invariance conditions

Times	Items	sample size	Mean level	Mean Shape	Variance level	Variance shape	Cov. level shape
3	5	100	-0.005	-0.016	0.115	0.319	0.280
		200	-0.010	0.003	0.124	0.297	0.267
		500	-0.003	0.002	0.054	0.174	0.156
		1000	-0.005	0.008	0.036	0.083	0.070
	10	100	-0.017	-0.010	0.117	0.369	0.307
		200	0.002	-0.009	0.080	0.281	0.244
		500	-0.010	-0.030	0.087	0.203	0.186
		1000	-0.001	0	0.040	0.131	0.115
	15	100	0.002	-0.012	0.112	0.377	0.316
		200	-0.023	-0.008	0.127	0.344	0.288
		500	-0.020	-0.004	0.104	0.262	0.224
		1000	0.008	-0.005	0.059	0.153	0.137
5	5	100	-0.018	-0.013	-0.020	-0.019	-0.021
		200	0.002	-0.004	-0.008	-0.017	-0.013
		500	0.006	-0.004	-0.011	-0.011	-0.006
		1000	-0.005	0.010	0.011	-0.004	0
	10	100	-0.013	-0.009	-0.022	-0.022	-0.026
		200	-0.006	-0.008	-0.012	-0.003	-0.011
		500	-0.012	-0.006	-0.003	-0.008	0.004
		1000	-0.006	-0.009	-0.007	-0.004	-0.004
	15	100	-0.017	-0.010	-0.023	-0.036	-0.010
		200	-0.007	-0.014	-0.005	-0.013	-0.003
		500	0.013	-0.001	-0.001	-0.003	-0.009
		1000	-0.008	-0.004	-0.018	0.008	-0.002

Note. Cov. = covariance

The results of the curve-of-factors model with respect to the relative bias of the chi-square statistic and the percentage of fit indices (i.e. CFI, TLI, RMSEA) indicating acceptable fit were almost identical regardless of whether the inadmissible solutions were removed (i.e. the difference in relative bias of chi-square was below 0.01 and the differences in percentage of CFI, TLI, and RMSEA indicating acceptable models was below 1%). Therefore, the results will be reported after removing the inadmissible solutions. The comparisons were based on a set of 1,000 admissible solutions for each condition.

The chi-square statistic obtained with the curve-of-factors model was affected by the sample size (partial $\eta^2 = 0.41$), the number of items (partial $\eta^2 = 0.14$), the number of measurement times (partial $\eta^2 = 0.06$), the interactions between the number of items and sample size (partial $\eta^2 = 0.17$), and the interaction between the number of measurement times and sample size (partial $\eta^2 = 0.08$), but not by reliability or type of item. The relative biases of the chi-square statistic by items, sample size, and measurement times are reported in Table 6. With sample sizes of 100 and 200, the chi-square statistics were positively biased, and Table 6 shows that the amount of bias increases as the number of items increases but decreases as sample size increases. The relative bias of the chi-square was positively biased by as much as 49% with 15 items and sample size of 100. Considering the relative bias of chi-square acceptable if it is smaller than 0.05, the relative bias of the chi-square was only consistently acceptable if the sample size was higher than 500, as shown in Table 6.

Table 6
Relative bias of the Chi-square statistic with the curve-of-factors model, collapsing across reliability, type of item, and invariance conditions

Items	Sample size	Relative bias of chi-square statistic	
		3 times	5 times
5	100	0.082	0.125
	200	0.038	0.058
	500	0.015	0.023
	1000	0.008	0.011
10	100	0.150	0.265
	200	0.069	0.110
	500	0.025	0.040
	1000	0.013	0.020
15	100	0.233	0.487
	200	0.100	0.172
	500	0.037	0.060
	1000	0.018	0.028

For each fit index, the relationship between each condition and whether the fit index was acceptable was evaluated using the Cramer's V coefficient. This coefficient ranges between zero and one, with one indicating a perfect relationship. Because of the large sample size, effects associated with Cramer's V as low as 0.014 were found to be statistically significant. Therefore, only Cramer's V coefficients larger than 0.1 will be reported.

The comparative fit index (CFI) depended on sample size (Cramer's V = 0.611), number of items (Cramer's V = 0.36), reliability (Cramer's V = 0.154), and number of measurement times (Cramer's V = 0.129).

The Tucker-Lewis index (TLI) depended on sample size (Cramer's V = 0.619), number of items (Cramer's V = 0.35), reliability (Cramer's V = 0.158), and number of measurement times (Cramer's V = 0.129).

The percentages of CFI and TLI that would suggest adequate fit of the model (i.e. CFI ≥ 0.95, TLI ≥ 0.95) (Hu & Bentler, 1999) for each condition were similar. The percentage of replications in which the CFI and TLI would suggest an adequate fit of the model increased as sample and reliability increased, but decreased as the number of items increased (see Table 7). With reliability equal to 0.7, five measurement times, sample size of 100 and 15 items, 0% of the models would be considered acceptable based on either the CFI or TLI. With sample sizes of 500 or 1000, at least 97% of the CFI and TLI would suggest the retention of the model, regardless of reliability, number of items and measurement times.

The root mean square error of approximation (RMSEA) depended on the sample size (Cramer's V = 0.569), number of items (Cramer's V = 0.194), and number of measurement times (Cramer's V = 0.149). Differently from the CFI and TLI, the RMSEA did not depend on the composite's reliability. The RMSEA tended to suggest the retention of the model more frequently with smaller sample sizes than the CFI or TLI. With a sample size of 200, at least 97.7% of the RMSEA were acceptable, regardless of the other conditions (see Table 7). However, with a sample size of 100, five measurement times and 15 items, 0% of the analyses produced values for CFI, TLI or RMSEA that would support the adequate fit of the model.

The fit of the models was also evaluated based on the combination of CFI, TLI and RMSEA suggested by Hu and Bentler (1999). These authors suggested that the fit of a model could be considered acceptable if CFI ≥ 0.95, TLI ≥ 0.95 and RMSEA ≤ 0.05. The fit of the model based on the combined criterion depended on sample size (Cramer's

V = 0.635), number of items (Cramer’s V = 0.323), reliability (Cramer’s V = 0.166), and number of measurement times (Cramer’s V = 0.122). The percentage of replications in which the fit of the model would be considered acceptable based on the combined criterion was at least 97% with sample sizes of 500 or 1000, regardless of the other conditions. On the other hand, the percentage of replications resulting in acceptable models according to the combined criterion was as low as 0% when the sample size was 100 and the number of items per construct was fifteen.

Table 7
Percentage of replications in which the fit indices would lead to model retention with the curve-of-factors model, collapsing across measurement times, type of item, and invariance conditions.

Times	Items	Sample	CFI		TLI		RMSEA		Combined	
			Reliability							
			0.7	0.9	0.7	0.9	0.7	0.9	0.7	0.9
3	5	100	84.9%	99.9%	79.3%	99.7%	85.5%	84.7%	79.0%	84.7%
		200	99.4%	100%	98.3%	100%	99.7%	99.7%	98.3%	99.7%
		500	100%	100%	100%	100%	100%	100%	100%	100%
		1000	100%	100%	100%	100%	100%	100%	100%	100%
	10	100	17.2%	86.2%	15.3%	81.7%	89.5%	88.1%	15.3%	81.2%
		200	86.6%	100%	82.8%	100%	100%	100%	82.8%	100%
		500	100%	100%	100%	100%	100%	100%	100%	100%
		1000	100%	100%	100%	100%	100%	100%	100%	100%
	15	100	0.1%	3.1%	0%	2.2%	62.0%	62.7%	0%	2.2%
		200	29.5%	99.4%	27.4%	99.0%	100%	100%	27.4%	99.0%
		500	99.6%	100%	99.4%	100%	100%	100%	99.4%	100%
		1000	100%	100%	100%	100%	100%	100%	100%	100%
5	5	100	67.4%	99.8%	62.6%	99.6%	90.4%	89.6%	62.6%	89.6%
		200	99.7%	100%	99.4%	100%	100%	100%	99.4%	100%
		500	100%	100%	100%	100%	100%	100%	100%	100%
		1000	100%	100%	100%	100%	100%	100%	100%	100%
	10	100	0%	2.2%	0%	1.7%	40%	39.7%	0%	1.7%
		200	43.3%	100%	39.3%	100%	100%	100%	39.3%	100%
		500	100%	100%	100%	100%	100%	100%	100%	100%
		1000	100%	100%	100%	100%	100%	100%	100%	100%
	15	100	0%	0%	0%	0%	0%	0%	0%	0%
		200	0%	39.7%	0%	34.6%	100%	100%	0%	34.6%
		500	98.2%	100%	97.6%	100%	100%	100%	97.6%	100%
		1000	100%	100%	100%	100%	100%	100%	100%	100%

Note. Criteria for model retention used to calculate the percentages:
CFI $\geq$ 0.95, TLI $\geq$ 0.95, RMSEA $\leq$ 0.05,
Combined = Criterion of acceptable fit based on the combination of CFI $\geq$ 0.95, TLI $\geq$ 0.95 and RMSEA $\leq$ 0.05 (Hu & Bentler, 1999)

Univariate LGM of item means

The analyses fitting the univariate latent growth model resulted in 100% of the models reaching convergence. However, like the curve-of-factors model, this model also produced a large number of non-positive variance/covariance matrices of the estimates. The number of inadmissible solutions was substantially larger with three measurement times than five measurement times (see Table 8). With three measurement times and 100 observations, the percentage of non-positive definite solutions was as high as 31%.

Table 8
Percentage of inadmissible solutions with the univariate latent growth model

Sample size	Times		Total
	3	5	
100	30.7%	2.4%	16.6%
200	23.3%	0.4%	11.9%
500	14.8%	0%	7.4%
1000	10.0%	0%	5.0%

Because removing the inadmissible solutions results in the truncation of the distributions of the estimates of the variances of level and shape and of the covariance between level and shape, the results with respect to relative bias will be presented both with and without removing the inadmissible solutions.

Without removing the inadmissible solutions, the univariate latent growth model produced positively biased estimates of the mean of the level in all conditions (see Table 9). For the estimates of the mean of the level, the relative bias depended on the type of item (partial $\eta^2 = 0.847$), invariance (partial $\eta^2 = 0.181$), number of measurement times (partial $\eta^2 = 0.069$), and the interaction between number of measurement times and invariance (partial η^2 = 0.089). The average bias of the estimate of mean of the level was positive and around 0.5 with tau-equivalent items (see Table 9) and 0.25 with congeneric items (see Table 9). The smaller bias resulting with the congeneric conditions as compared with the tau-equivalent conditions is because the loadings of the congeneric conditions are lower than one. These loadings weight less the estimates of the mean of the level than loadings equal one.

The relative bias of the mean of the level that has been found in this study is a consequence of summing items with different intercepts to form each composite. The mean μ_{y_i} of each item y_i at the first measurement time is the sum of the mean of the latent factor μ_η multiplied by the item's factor loading λ_{y_i}, and the item's intercept τ_{y_i} :

$$\mu_{y_i} = \lambda_{y_i}\mu_\eta + \tau_{y_i} \tag{54}$$

While the curve-of-factors model evaluates the growth of the latent factor, the three univariate latent growth models evaluated in this study are designed to analyze the composite *c*. The mean of the composite, μ_c, formed by *k* items at the first measurement time is:

$$\mu_c = \frac{\sum_{i=1}^{k} \mu_{y_i}}{k} \tag{55}$$

Given that $\mu_{y_i} = \lambda_y \mu_\eta + \tau_{y_i}$ (56)

Then $\mu_c = \frac{\sum_{i=1}^{k} (\lambda_i \mu_\eta + \tau_{y_i})}{k}$ and $\mu_c = \frac{\sum_{i=1}^{k} \lambda_i \mu_\eta}{k} + \frac{\sum_{i=1}^{k} \tau_{y_i}}{k}$. (57)

Defining the average of the factor loadings as $\mu_\lambda = \frac{\sum_{i=1}^{k} \lambda_i}{k}$ and the intercepts as $\mu_\tau = \frac{\sum_{i=1}^{k} \tau_{y_i}}{k}$

Then $\mu_c = \mu_\lambda \mu_\eta + \mu_\tau$. (58)

Considering that the mean of the composites μ_c is the mean of the latent factor μ_η weighted by the average of the factor loadings μ_λ plus the average of the intercepts μ_τ, the mean of the level estimated with univariate latent growth models will be different from the mean of the level of the latent factor, unless the loadings sum to one and the intercepts are zero. Because the scale of measurement of the composites is arbitrary, this shift of the mean of the level may not affect the interpretation of the results. This difference between the estimate of the mean of the level and the population mean of the level will not change the interpretation of the results if a single group is being studied, unless the researcher attempts to interpret the mean of the shape as the percentage of change with respect to the mean of the level. For example, in this study the population values of the mean of the level and mean of the shape were set to 1. These values indicate that, between two measurement times, individuals present an average growth equal to 100% of the mean of level. However, if the mean of the level is estimated to be 1.5, which corresponds to a relative bias of 0.5, the interpretation of the percentage of growth with respect to the mean of the level would change to 66%. If the mean of the level is used to compare two groups with respect to their status at a given time, both groups will have their estimated mean of the level shifted by the same amount and the difference between groups will remain the same if there is strict factorial invariance. However, if the items have different loadings and/or intercepts across groups, the estimated mean of the level of the groups will be shifted by different amounts from their population means, resulting in different estimates of the mean of the level even if the groups have the same population mean of the level.

The relative bias of the estimates of the mean of the shape with the univariate latent growth model depended on the type of item (partial $\eta^2 = 0.944$), invariance (partial $\eta^2 = 0.646$), number of measurement times (partial $\eta^2 = 0.342$), number of items (partial $\eta^2 = 0.322$), the interaction between number of measurement times and invariance (partial $\eta^2 = 0.327$), the interaction between the number of items and invariance (partial $\eta^2 = 0.212$), and the interaction between number of items and number of measurement times (partial $\eta^2 = 0.112$). The univariate latent growth model yielded unbiased estimates of the mean of the shape when the items were essentially tau-equivalent and there was either strict factorial invariance or weak factorial invariance with five measurement times (see Table 9). With tau-equivalent items, weak factorial invariance, and three measurement times, the estimates of the mean of the shape were negatively biased by as much as 19%. The estimates of the mean of the shape were negatively biased by as much as 49% with congeneric items, weak factorial invariance, three measurement times and sample size of 100. The overall mean relative bias of the mean of the shape with essentially congeneric items was -0.288.

The bias of the estimates of the variances of the level and shape depended only on the type of item (partial η^2 = 0.674 for the estimate of the variance of the level and partial η^2 = 0.252 for the estimate of the variance of the shape). The estimates of the variances of the level and shape had acceptable bias in the conditions with essentially tau-equivalent items and large negative biases in the conditions with essentially congeneric items (see Table 9). The mean relative bias of the variances of the level and shape for the conditions with essentially congeneric items, collapsing across all other conditions, was -0.48.

The relative bias of the estimate of the covariance between level and shape depended on the type of item (partial η^2 = 0.208) and on invariance (partial η^2 = 0.059). The estimate of the covariance between level and shape was unbiased if the items were tau-equivalent and the number of measurement times was three (see Table 9). With five measurement times, the estimate of the covariance also had acceptable bias when the items were tau-equivalent and there was strict factorial invariance. With five measurement times and weak factorial invariance, the covariance was unbiased if the number of items per construct was either 10 or 15 and the sample size was higher than 100. The estimate of the covariance between level and shape had negative bias when the items were essentially congeneric (see Table 9). The estimate of the covariance between level and shape was negatively biased as much as 51% with essentially congeneric items, three times, five items, and strict factorial invariance. The mean relative bias of the estimate of the covariance between level and shape with essentially congeneric items, collapsing across all other conditions, was -0.363.

The univariate latent growth model produced unbiased estimates of the standard errors in all conditions, as shown in Table 10. The relative bias of the standard errors did not show dependence on any of the conditions included in the study.

Table 9
Mean relative bias of parameter estimates across conditions with the univariate latent growth model, without removing the inadmissible solutions, collapsing across sample size and reliability.

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. Shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.560	-0.001	-0.011	-0.034	0.023
		10	0.520	0.000	-0.005	-0.010	0.005
		15	0.534	0.000	-0.011	-0.027	0.021
	5	5	0.560	0.000	-0.007	-0.006	0.003
		10	0.520	0.000	-0.007	-0.009	0.002
		15	0.533	0.000	-0.009	-0.012	-0.002
Weak	3	5	0.545	-0.193	-0.014	-0.034	0.028
		10	0.512	-0.101	-0.011	-0.014	0.010
		15	0.517	-0.065	-0.012	-0.024	0.024
	5	5	0.432	-0.047	-0.050	-0.059	0.110
		10	0.443	-0.013	-0.024	-0.028	0.043
		15	0.482	-0.016	-0.015	-0.014	0.017
Congeneric							
Strict	3	5	0.260	-0.300	-0.513	-0.513	-0.510
		10	0.270	-0.250	-0.442	-0.445	-0.438
		15	0.259	-0.273	-0.477	-0.481	-0.472
	5	5	0.260	-0.300	-0.514	-0.513	-0.508
		10	0.271	-0.250	-0.439	-0.442	-0.437
		15	0.260	-0.273	-0.475	-0.475	-0.471
Weak	3	5	0.246	-0.492	-0.516	-0.524	-0.500
		10	0.260	-0.351	-0.441	-0.446	-0.424
		15	0.243	-0.339	-0.479	-0.491	-0.460
	5	5	0.131	-0.349	-0.553	-0.562	-0.418
		10	0.192	-0.264	-0.456	-0.461	-0.399
		15	0.209	-0.290	-0.484	-0.482	-0.456
Configural	3	5	0.243	-0.316	-0.502	-0.219	-0.225
		10	0.202	-0.265	-0.583	-0.818	0.207
		15	0.217	-0.230	-0.539	-0.548	-0.058
	5	5	0.194	-0.249	-0.477	-0.376	-0.308
		10	0.166	-0.180	-0.486	-0.333	-0.283
		15	0.233	-0.228	-0.457	-0.383	-0.366

Note. Tau Eq. = Essentially tau-equivalent, Congeneric = Essentially congeneric, Var. = Variance, Cov. = Covariance

Table 10

Mean relative bias of the estimates of standard errors across conditions with the univariate latent growth model, without removing the inadmissible solutions, collapsing across sample size and reliability.

Type of item Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. Shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.007	-0.005	0.001	-0.006	-0.002
		10	-0.003	-0.002	-0.016	-0.009	-0.018
		15	-0.003	-0.003	-0.016	-0.001	-0.006
	5	5	-0.001	-0.002	-0.005	-0.026	-0.012
		10	-0.012	-0.012	-0.015	-0.014	-0.006
		15	-0.001	-0.012	0.014	0	-0.005
Weak	3	5	0.003	-0.003	-0.002	-0.016	-0.007
		10	-0.008	-0.007	-0.016	-0.011	-0.012
		15	-0.002	-0.008	-0.001	-0.001	-0.006
	5	5	-0.023	-0.011	0.002	0.006	0.007
		10	-0.005	-0.005	0	0.008	0.006
		15	-0.015	-0.012	-0.006	0.012	0.006
Congeneric							
Strict	3	5	-0.010	-0.017	0.001	-0.017	-0.017
		10	-0.013	-0.009	-0.005	-0.006	-0.007
		15	-0.010	-0.016	-0.007	-0.017	0.001
	5	5	0.002	0.002	0.002	-0.014	0.004
		10	0.001	-0.005	-0.013	-0.004	-0.008
		15	-0.010	0.007	-0.012	-0.013	-0.005
Weak	3	5	-0.010	-0.017	-0.008	-0.001	-0.003
		10	-0.012	-0.021	-0.006	-0.012	-0.006
		15	-0.004	-0.015	-0.020	-0.022	-0.019
	5	5	-0.069	-0.085	0.005	-0.016	0.004
		10	-0.034	-0.035	-0.016	0.006	0.010
		15	-0.011	0	-0.009	-0.015	-0.016
Configural	3	5	-0.021	-0.006	-0.006	-0.009	-0.010
		10	-0.055	-0.052	-0.018	-0.014	-0.031
		15	-0.042	-0.038	-0.011	-0.027	-0.015
	5	5	-0.023	-0.051	0.017	0.003	-0.025
		10	-0.043	-0.026	-0.013	-0.003	-0.013
		15	-0.021	-0.025	-0.007	0.003	-0.008

Note. Tau Eq. = Essentially tau-equivalent, Congeneric = Essentially congeneric, Var. = Variance, Cov. = Covariance

The results obtained from the univariate latent growth model removing the inadmissible solutions indicate that this model results in positive relative bias of the estimates of the mean of the level in all conditions (see Table 11). The relative bias of the estimate of the mean of the level depended on type of item (partial η^2 = 0.846), invariance (partial η^2 = 0.174) and the number of measurement times (partial η^2 = 0.078). The overall mean of the relative bias of the estimate of the mean of the level for conditions with tau-equivalent items was 0.5. For the conditions with congeneric items, the mean relative bias was 0.2.

The relative bias of the estimate of the mean of the shape depended on the type of item (partial η^2 = 0.944), invariance (partial η^2 = 0.646), number of measurement times (partial η^2 = 0.340), number of items (partial η^2 = 0.322), the interaction between the number of measurement times and invariance (partial η^2 = 0.327), the interaction between the number of items and invariance (partial η^2 = 0.212), the interaction between the number of items and the number of measurement times (partial η^2 = 0.111), and the interaction between type of item and number of items (partial η^2 = 0.086). The bias of the estimate of the mean of the shape was acceptable in the conditions with essentially tau-equivalent items, with the exception of the conditions with weak factorial invariance and three measurement times, where the bias was negative (see Table 11). For the conditions with essentially tau-equivalent items, the mean relative bias of the estimate of the mean of the shape was -0.036. The relative bias of the estimate of the mean of the shape was unacceptable and negative in all conditions with essentially congeneric items (see Table

11). The relative biased tended to decrease as the number of items increased. For the conditions with congeneric items, the mean relative bias was -0.289.

The relative bias of the estimate of the variance of the level depended on the type of item (partial $\eta^2 = 0.722$). The relative bias of the estimate of the variance of the level were acceptable on the conditions with essentially tau-equivalent items but unacceptable on the conditions with essentially congeneric items (see table 11). The overall mean of the relative bias of the estimate of the variance of the level for the conditions with essentially congeneric items was -0.47. The relative bias for the conditions with essentially congeneric items was consistent across reliability, number of measurement times, number of items, and invariance conditions.

The relative bias of the estimate of the variance of the shape depended on type of item (partial $\eta^2 = 0.411$), and the number of measurement times (partial $\eta^2 = 0.05$). The relative bias was only acceptable in conditions with essentially tau-equivalent items and five measurement times. The conditions with essentially tau-equivalent items and three measurement times presented positive bias consistently around 0.17. The conditions with essentially congeneric items presented negative bias. The bias was consistent across all the other conditions. The overall mean of the relative bias of the estimate of the variance of the shape with essentially congeneric items was -0.397.

With respect to the relative bias of the estimates of the correlation between level and shape, none of the conditions had an effect size higher than 0.05. Although the ANOVA results indicated that there is no practical difference between the conditions included in this study with respect to the relative bias of the estimates of the correlation, most of the estimates presented unacceptable relative biases. The relative bias of the correlation was only consistently acceptable in the condition with strict factorial invariance and five measurement times (see Table 11). In the conditions with weak factorial invariance and five measurement times, the bias was only acceptable when the number of items was fifteen. In the conditions with three measurement times, the bias of the correlation tended to be acceptable in conditions where the sample size was 500 or above, with either strict or weak factorial invariance. Conditions with strict factorial invariance and three measurement times presented consistent negative bias, around -0.12. Conditions with configural invariance always presented positive bias of the estimate of the correlation between level and shape, regardless of the sample size and number of measurement times. The estimate of the correlation between level and shape was positively biased by as much as 55% with configural invariance, essentially congeneric items, 3 measurement times and 15 items per construct.

Table 11
Mean relative bias of parameter estimates across conditions with the univariate latent growth model, after removing the inadmissible solutions, collapsing across sample size and reliability.

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. Shape	Cor. level shape
Tau Eq.							
Strict	3	5	0.561	0	0.041	0.169	-0.113
		10	0.519	0	0.042	0.178	-0.120
		15	0.534	0	0.043	0.178	-0.121
	5	5	0.560	0	-0.005	-0.004	0.031
		10	0.520	0	-0.005	-0.007	0.032
		15	0.533	0	-0.007	-0.010	0.028
Weak	3	5	0.545	-0.192	0.040	0.176	-0.127
		10	0.513	-0.102	0.039	0.173	-0.110
		15	0.517	-0.065	0.039	0.176	-0.109
	5	5	0.432	-0.047	-0.047	-0.056	0.193
		10	0.442	-0.013	-0.021	-0.026	0.088
		15	0.482	-0.016	-0.013	-0.012	0.052
Congeneric							
Strict	3	5	0.261	-0.300	-0.491	-0.422	-0.111
		10	0.270	-0.250	-0.415	-0.339	-0.116
		15	0.260	-0.274	-0.451	-0.379	-0.123
	5	5	0.260	-0.300	-0.513	-0.512	0.033
		10	0.270	-0.250	-0.438	-0.441	0.027
		15	0.260	-0.273	-0.474	-0.474	0.029
Weak	3	5	0.247	-0.492	-0.492	-0.425	-0.111
		10	0.260	-0.351	-0.414	-0.333	-0.121
		15	0.245	-0.339	-0.451	-0.383	-0.112
	5	5	0.131	-0.349	-0.551	-0.559	0.329
		10	0.193	-0.264	-0.455	-0.460	0.130
		15	0.209	-0.290	-0.483	-0.481	0.072
Configural	3	5	0.178	-0.301	-0.496	-0.500	0.177
		10	0.244	-0.316	-0.472	-0.110	0.084
		15	0.213	-0.264	-0.491	-0.415	0.558
	5	5	0.222	-0.229	-0.485	-0.330	0.362
		10	0.194	-0.249	-0.475	-0.374	0.224
		15	0.167	-0.180	-0.484	-0.331	0.239

Note. Tau-eq. = Essentially tau-equivalent, Congeneric = Essentially congeneric, Var. = variance, Cor. = correlation

The relative biases of the estimated standard errors of the means of the level and shape did not depend on any of the conditions included in the study (see Table 12). Furthermore, the relative bias of the estimate of the standard error of the mean of the level was acceptable in all of the conditions. The relative bias of the estimated standard error of the mean of the shape was acceptable in the conditions with either strict or weak factorial invariance. In the condition with configural invariance and five measurement times, the relative bias of the estimated standard error of the mean of the shape was also acceptable. However, in the conditions with configural invariance and three measurement times, the bias was unacceptable and positive, with a magnitude as large as 0.27.

The relative bias of the estimate of the standard error of variance of the level depended on the number of measurement times (partial η^2 = 0.249). The relative bias of the estimate of the standard error of the variance of the level was marginally acceptable in the conditions with essentially tau-equivalent items (see Table 12). In the conditions with essentially congeneric items, the bias of the estimate of the standard error of the variance of the level was acceptable when the number of measurement times was five (see Table 12). With essentially congeneric items and three measurement times, the bias of the estimate of the standard error of the variance of the level was positive. The estimate of the standard error of the variance of the level was positively biased by as much as 67% in conditions with essentially congeneric items, configural invariance, three measurement times and 10 items.

The relative bias of the estimate of the standard error of the variance of the shape depended on number of measurement times (partial η^2 = 0.578), sample size (partial η^2 =

0.85), number of items (partial η^2 = 0.077), the interaction between number of measurement times and sample size (partial η^2 = 0.123), the interaction between number of items and invariance (partial η^2 = 0.109) and the interaction between number of measurement times and invariance (partial η^2 = 0.108). The relative bias of the estimate of the standard error of the variance of the shape only had acceptable bias when the number of measurement times was five. When the number of measurement times was three, the estimate of the standard error of the variance of the shape was positively biased under all conditions. The estimate was biased by around 20%, however the bias was as large as 65% in conditions with essentially congeneric items, configural invariance, three measurement times and 10 items.

The relative bias of the estimate of the standard error of the covariance between the level and shape depended on the number of measurement times (partial η^2 = 0.609), invariance (partial η^2 = 0.143), sample size (partial η^2 = 0.102), number of items (partial η^2 = 0.083), reliability (partial η^2 = 0.077), the interaction between number of times and invariance (partial η^2 = 0.178), the interaction between number of items and invariance (partial η^2 = 0.133), and the interaction between number of times and sample size (partial η^2 = 0.124). The estimate had acceptable bias when the items were essentially tau-equivalent and the number of measurement times was five. Additionally, the estimate had acceptable bias in all conditions with essentially congeneric items (see Table 12). When the items were essentially tau-equivalent and the number of measurement times was three, the relative bias was positive and the magnitude was around 0.18 (see Table 12).

Table 12
Mean relative bias of the estimates of standard errors across conditions with the univariate latent growth model, after removing the inadmissible solutions, collapsing across sample size and reliability.

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.008	-0.006	0.104	0.203	0.189
		10	-0.003	0.012	0.082	0.199	0.174
		15	0.000	0.002	0.076	0.200	0.189
	5	5	-0.001	-0.001	0.003	-0.020	-0.002
		10	-0.011	-0.011	-0.007	-0.009	0.004
		15	-0.001	-0.011	0.022	0.003	0.002
Weak	3	5	-0.023	-0.005	0.104	0.217	0.209
		10	-0.007	-0.001	0.098	0.208	0.190
		15	-0.016	-0.002	0.085	0.216	0.182
	5	5	-0.044	-0.032	0.009	0.004	0.018
		10	-0.017	-0.033	0.012	0.002	0.000
		15	-0.006	-0.012	-0.008	0.001	0.005
Congeneric							
Strict	3	5	-0.002	0.089	0.191	0.161	-0.002
		10	-0.005	0.082	0.200	0.178	-0.005
		15	-0.009	0.083	0.191	0.181	-0.009
	5	5	0.003	0.008	-0.009	0.010	0.003
		10	-0.003	-0.007	0.003	0.003	-0.003
		15	0.006	-0.007	-0.007	0.005	0.006
Weak	3	5	-0.015	0.075	0.230	0.193	-0.015
		10	-0.016	0.090	0.211	0.200	-0.016
		15	-0.013	0.096	0.201	0.196	-0.013
	5	5	-0.082	0.019	0.003	0.030	-0.082
		10	-0.035	-0.010	0.013	0.020	-0.035
		15	0.000	-0.004	-0.009	-0.008	0.000
Configural	3	5	-0.009	0.109	0.173	0.179	-0.009
		10	-0.034	0.273	0.674	0.650	-0.034
		15	-0.038	0.187	0.390	0.384	-0.038
	5	5	-0.050	0.028	0.010	-0.015	-0.050
		10	-0.025	-0.002	0.004	0.000	-0.025
		15	-0.023	-0.001	0.009	-0.001	-0.023

Note. Tau-eq. = Essentially tau-equivalent, Congeneric = Essentially congeneric, Var. = variance, Cor. = correlation

Because the relative bias of the chi-square statistic was similar regardless of whether the inadmissible solutions were removed or not, the analyses of the chi-square statistic and fit indices were performed after removing the inadmissible solutions. The relative bias of the chi-square statistic depended on invariance (partial $\eta^2 = 0.335$), sample size (partial $\eta^2 = 0.285$), number of measurement times (partial $\eta^2 = 0.069$), reliability (partial $\eta^2 = 0.062$), the interaction between sample size and invariance (partial $\eta^2 = 0.235$), the interaction between number of measurement times and invariance (partial $\eta^2 = 0.192$), the interaction between number of items and invariance (partial $\eta^2 = 0.171$) and the interaction between number of items and number of times (partial $\eta^2 = 0.078$).

The relative bias of the chi-square statistic was acceptable in conditions with strict factorial invariance (see Table 13), regardless of sample size, reliability, number of measurement times, type of item and number of items. With either weak or configural invariance, the relative bias of the chi-square statistic was positive and increased as sample size increased. Table 14 displays the average chi-square statistic grouped according to invariance, reliability, number of measurement times and sample size. In Table 14, it possible to notice that the chi-square values stay close to the degrees of freedom when there is strict factorial invariance, but become larger than the degrees of freedom when there is weak or configural invariance.

Table 13
Relative bias of the chi-square with the univariate latent growth model, collapsing across type of item and number of items.

Reliability	Times	Sample	Invariance		
			Strict	Weak	Configural
0.70	3	100	0.0195	0.380	3.584
		200	0.014	0.728	7.104
		500	0.002	1.955	18.072
		1000	0.011	3.809	35.936
	5	100	0.024	0.922	1.106
		200	0.007	1.799	2.171
		500	-0.001	4.459	5.422
		1000	0.003	8.902	10.797
0.90	3	100	-0.005	0.744	6.656
		200	-0.015	1.461	13.272
		500	-0.003	3.727	33.567
		1000	0.009	7.320	67.847
	5	100	0.029	1.851	2.262
		200	0.017	3.640	4.485
		500	0.010	9.058	11.072
		1000	0.004	18.122	22.187

Table 14
Average Chi-square statistics with the univariate latent growth model, collapsing across type of item and number of items.

Reliability	Times	Sample	Invariance		
			Strict	Weak	Configural
0.70	3	100	1.019	1.380	4.584
		200	1.014	1.728	8.104
		500	1.002	2.955	19.072
		1000	1.011	4.809	36.936
	5	100	10.243	19.215	21.059
		200	10.072	27.988	31.713
		500	9.991	54.588	64.218
		1000	10.031	99.022	117.972
0.90	3	100	0.995	1.744	7.656
		200	0.985	2.461	14.272
		500	0.997	4.727	34.567
		1000	1.009	8.320	68.847
	5	100	10.293	28.513	32.616
		200	10.169	46.399	54.846
		500	10.102	100.585	120.723
		1000	10.044	191.219	231.865

Note. df = 1 for 3 measurement times, df = 10 for 5 measurement times

With the univariate latent growth model, the CFI depended on invariance (Cramer's V = 0.452), type of item (Cramer's V = 0.211), number of measurement times (Cramer's V = 0.185), and number of items (Cramer's V = 0.136). The TLI depended on invariance (Cramer's V = 0.521), type of item (Cramer's V = 0.249), and number of items (Cramer's V = 0.103). The percentages of CFI and TLI which would indicate the adequate fit of the model were similar. If there was strict factorial invariance, at least 99.6% of the CFI and 96.6% of the TLI would indicate adequate model fit, regardless of the other conditions. However, with weak or strict factorial invariance, the percentage of CFI and TLI equal or above 0.95 depended on the combination of the other conditions (see Table 15). With weak factorial invariance, essentially tau-equivalent items, five measurement times and five items, only 33% of the CFI and TLI were equal or above 0.95. With weak factorial invariance, essentially congeneric items, five measurement times and five items, the proportion of CFI and TLI equal or above 0.95 was 1.6%. In the conditions with configural invariance and congeneric items, the percentages of CFI and TLI equal or above 0.95 tended to be lower than in the other conditions (see Table 15).

The RMSEA depended on invariance (Cramer's V = 0.630), type of item (Cramer's V = 234), and number of times (Cramer's V = 0.119). The percentage of RMSEA which would indicate an acceptable model was consistently lower than the CFI and TLI in all conditions (see Table 15). With strict factorial invariance, the percentage of RMSEA equal or below 0.05 was between 82.7% and 90% for all conditions. With weak factorial invariance, the percentage of RMSEA equal or below 0.05 was as low as 1.2% with essentially tau-equivalent items, five measurement times and five items. With configural invariance, the percentage of RMSEA equal or below 0.05 was 25.5% with congeneric items, three measurement times and 3 items, but was below 4.4% with all other conditions (see Table 15).

The fit of the model based on the combined criterion of CFI ≥ 0.95, TLI ≥ 0.95 and RMSEA ≤ 0.05 depended on invariance (Cramer's V = 0.630), type of item (Cramer's V = 0.234), and number of measurement times (Cramer's V = 0.119). Because the RMSEA would indicate the lowest percentage of acceptable models, the percentage of models that met the combined criterion for acceptable fit was identical to the percentage of models that had a RMSEA equal or below 0.05 for all conditions (see Table 15).

It is interesting to notice that, differently from the curve-of-factors model, the CFI, TLI and RMSEA used to evaluate the fit of the univariate latent growth model were not affected by sample size (see Tables 7 and 36).

Table 15
Percentage of replications in which the fit indices would lead to model retention with the univariate latent growth model, collapsing across sample size and reliability conditions.

Invariance	Type of item	Times	Items	CFI	TLI	RMSEA	Combined
Strict	Tau Eq.	3	5	99.6%	96.5%	82.5%	82.5%
			10	99.6%	96.6%	82.9%	82.9%
			15	99.6%	97.0%	83.2%	83.2%
		5	5	99.8%	99.8%	89.7%	89.7%
			10	99.7%	99.7%	88.8%	88.8%
			15	99.8%	99.8%	89.8%	89.8%
	Congeneric	3	5	99.6%	96.8%	83.3%	83.3%
			10	99.7%	96.6%	82.7%	82.7%
			15	99.7%	97.2%	82.7%	82.7%
		5	5	99.9%	99.9%	89.5%	89.5%
			10	99.8%	99.8%	89.7%	89.7%
			15	99.8%	99.8%	90.0%	90.0%
Weak	Tau Eq.	3	5	98.9%	91.8%	54.8%	54.8%
			10	99.4%	95.3%	72.4%	72.4%
			15	98.6%	90.4%	49.3%	49.3%
		5	5	33.4%	33.4%	1.2%	1.2%
			10	96.6%	96.6%	14.7%	14.7%
			15	99.2%	99.2%	61.1%	61.1%
	Congeneric	3	5	98.0%	84.1%	37.2%	37.2%
			10	99.4%	93.8%	63.4%	63.4%
			15	97.9%	80.2%	32.4%	32.4%
		5	5	1.6%	1.6%	0.1%	0.1%
			10	82.4%	82.4%	4.2%	4.2%
			15	98.4%	98.4%	34.0%	34.0%
Configural	Congeneric	3	5	96.7%	72.2%	25.5%	25.5%
			10	20.0%	3.5%	1.2%	1.2%
			15	58.2%	11.6%	4.4%	4.4%
		5	5	53.8%	53.8%	2.5%	2.5%
			10	47.5%	47.5%	1.8%	1.8%
			15	51.7%	51.7%	2.0%	2.0%

Note. Criteria for model retention used to calculate the percentages:
CFI $\geq$ 0.95, TLI $\geq$ 0.95, RMSEA $\leq$ 0.05,
Combined = Criterion of acceptable fit based on the combination of CFI $\geq$ 0.95, TLI $\geq$ 0.95 and RMSEA $\leq$ 0.05 (Hu & Bentler, 1999)

Univariate LGM of item means with fixed error variances

Non-covergence was not a problem for the univariate LGM model with fixed error variances; 100% of the replications reached convergence. Despite this excellent convergence rate, a large number of inadmissible solutions occurred. The number of inadmissible solutions decreased as sample size and number of measurement times increased (see Table 16).

Table 16
Percentage of inadmissible solutions with the univariate latent growth model with fixed error variances

Sample size	Times		Total
	3	5	
100	31.3%	2.4%	16.9%
200	23.7%	0.4%	12.1%
500	14.9%	0%	7.4%
1000	10.0%	0%	5.0%

Without removing the inadmissible solutions, the relative bias of the estimate of the mean of the level obtained the univariate LGM with fixed error variances depended on type of item (partial η^2 = 0.847), invariance (partial η^2 = 0.181), number of measurement times (partial η^2 = 0.069), and interaction between number of measurement times and invariance (partial η^2 = 0.089). The estimate of the mean of the level was positively biased with all conditions (see Table 17). With essentially tau-equivalent items, the overall mean relative bias was 0.51. With essentially congeneric items, the overall mean relative bias of the mean of the level was 0.22. With both essentially tau-equivalent and essentially congeneric items, the magnitude of the relative bias of the estimates of the mean of the level was consistent across all conditions.

The relative bias of the estimate of the mean of the shape depended on type of item (partial η^2 = 0.944), invariance (partial η^2 = 0.646), number of measurement times (partial η^2 = 0.342), number of items (partial η^2 = 0.322), the interaction between number of times and invariance (partial η^2 = 0.327), the interaction between number of items and invariance (partial η^2 = 0.212), the interaction between number of items and number of times (partial η^2 = 0.112), and the interaction between number of items and type of items (partial η^2 = 0.084). The estimate of the mean of the shape was unbiased with essentially tau-equivalent items and strict factorial invariance. With tau-equivalent items and weak factorial invariance, the estimate was unbiased if the number of measurement times was five and negatively biased if the number of measurement times was three (see table 20). The overall mean relative bias of the mean of the shape with essentially tau-equivalent items was -0.036. The relative bias of the mean of the shape was as large -0.18 with tau-equivalent items, weak factorial invariance, 3 measurement times, and 5 items. The estimate of the mean of the shape was also negatively biased with congeneric items (see Table 21), and the overall mean relative bias with congeneric items was -0.289. The relative bias of the mean of the shape was as large as -0.492 with congeneric items, weak factorial invariance, three measurement times and five items.

The relative bias of the estimate of the variance of the level depended on the type of item (partial η^2 = 0.674). The relative bias was acceptable when the items were essentially tau-equivalent, regardless of the other conditions (see Table 17). The overall mean relative biase with essentially tau-equivalent items was -0.023. On the other hand, the estimate of the variance of the level was negatively biases when the items were essentially congeneric with all other conditions (see Table 17). The relative bias with essentially congeneric items was consistent across all conditions and their overall mean was -0.491.

The relative bias of the estimate of the variance of the shape depended only on type of item (partial η^2 = 0.252). This relative bias was acceptable with almost all conditions where the items were essentially tau-equivalent (see Table 17), and the overall mean relative bias was -0.022. The only condition where the bias was just marginally acceptable was with essentially tau-equivalent items, weak factorial invariance, five measurement times and five items (see Table 17). However, when the items were essentially congeneric, the univariate LGM with fixed error variances resulted in unacceptable and negative relative bias of the estimate of the variance of the shape in all conditions (see Table 21). The overall mean relative bias of the estimate of the variance of the shape with essentially congeneric items was -0.47. The relative bias was as high as -0.818 with essentially congeneric items, configural invariance, three measurement times and 10 items.

The relative bias of the estimate of the covariance between level and shape depended on type of item (partial η^2 = 0.208) and invariance (partial η^2 = 0.059). The univariate LGM with the fixed error variance produced an unbiased estimate of the covariance between level and shape when the items were essentially tau-equivelent (see

Table 20). The overall mean relative bias with essentially tau-equivalent items was 0.023. On the other hand, the estimate of the covariance between level and shape was negatively biased with essentially congeneric items, as shown in Table 17. The overall mean relative bias with essentially congeneric items was -0.363. The relative bias of the estimate of the covariance between level and shape was higher with essentially congeneric items and either strict or weak factorial invariance (i.e. its magnitude was between -0.399 and -0.511) than with essentially congeneric items and configural invariance (i.e. its magnitude was between -0.058 and -0.366).

None of the estimates of the standard errors obtained with the univariate LGM with fixed error variances showed a relationship with the conditions manipulated in this study. Furthermore, all of the relative biases of the estimate of the standard error calculated without removing the inadmissible solutions were acceptable (see Table 18).

Table 17
Mean relative bias of parameter estimates across conditions with the univariate LGM with fixed error variances, without removing the inadmissible solutions, collapsing across sample size and reliability

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. Shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.560	-0.001	-0.011	-0.034	0.023
		10	0.520	0.000	-0.005	-0.009	0.005
		15	0.534	0.000	-0.011	-0.027	0.021
	5	5	0.560	0.000	-0.007	-0.006	0.003
		10	0.520	0.000	-0.007	-0.009	0.002
		15	0.533	0.000	-0.009	-0.012	-0.002
Weak	3	5	0.545	-0.193	-0.014	-0.034	0.028
		10	0.512	-0.101	-0.011	-0.014	0.010
		15	0.517	-0.065	-0.012	-0.024	0.024
	5	5	0.432	-0.047	-0.050	-0.059	0.110
		10	0.443	-0.013	-0.024	-0.028	0.043
		15	0.482	-0.016	-0.015	-0.014	0.017
Congeneric							
Strict	3	5	0.260	-0.300	-0.513	-0.513	-0.511
		10	0.270	-0.250	-0.442	-0.445	-0.438
		15	0.259	-0.273	-0.477	-0.481	-0.472
	5	5	0.260	-0.300	-0.514	-0.513	-0.508
		10	0.271	-0.250	-0.439	-0.442	-0.437
		15	0.260	-0.273	-0.475	-0.475	-0.471
Weak	3	5	0.246	-0.492	-0.516	-0.524	-0.500
		10	0.260	-0.351	-0.441	-0.446	-0.424
		15	0.243	-0.339	-0.479	-0.491	-0.460
	5	5	0.131	-0.349	-0.553	-0.562	-0.418
		10	0.192	-0.264	-0.456	-0.461	-0.399
		15	0.209	-0.290	-0.484	-0.482	-0.456
Configural	3	5	0.243	-0.316	-0.502	-0.219	-0.225
		10	0.202	-0.265	-0.583	-0.818	0.207
		15	0.217	-0.230	-0.539	-0.548	-0.058
	5	5	0.194	-0.249	-0.477	-0.376	-0.308
		10	0.166	-0.180	-0.486	-0.333	-0.283
		15	0.233	-0.228	-0.457	-0.383	-0.366

Note. Tau Eq. = Essentially tau-equivalent, Congeneric = Essentially congeneric items, Var. = variance, Cov. = covariance

Table 18
Mean relative bias of the estimates of standard errors across conditions with the univariate LGM with fixed error variances, without removing the inadmissible solutions, collapsing across sample size and reliability

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.007	-0.005	0.001	-0.006	-0.003
		10	-0.004	-0.002	-0.016	-0.009	-0.018
		15	-0.003	-0.003	-0.016	-0.001	-0.006
	5	5	-0.001	-0.002	-0.005	-0.025	-0.011
		10	-0.012	-0.012	-0.015	-0.012	-0.006
		15	-0.001	-0.012	0.014	0.002	-0.004
Weak	3	5	-0.023	-0.011	0.002	0.006	0.007
		10	-0.006	-0.005	-0.001	0.008	0.006
		15	-0.015	-0.012	-0.007	0.012	0.006
	5	5	-0.045	-0.031	0.005	-0.005	0.004
		10	-0.016	-0.036	0.007	-0.001	-0.011
		15	-0.008	-0.014	-0.013	-0.001	-0.002
Congeneric							
Strict	3	5	-0.011	-0.017	0.001	-0.016	-0.017
		10	-0.014	-0.010	-0.005	-0.006	-0.007
		15	-0.010	-0.015	-0.007	-0.017	0.001
	5	5	0.003	0.003	0.002	-0.011	0.000
		10	0.001	-0.008	-0.012	-0.001	-0.014
		15	-0.010	0.010	-0.012	0.001	0.002
Weak	3	5	-0.010	-0.017	-0.008	-0.001	-0.002
		10	-0.012	-0.022	-0.006	-0.011	-0.006
		15	-0.004	-0.014	-0.020	-0.022	-0.019
	5	5	-0.069	-0.081	0.005	-0.005	0.004
		10	-0.034	-0.037	-0.016	0.001	0.004
		15	-0.011	-0.001	-0.009	-0.005	-0.009
Configural	3	5	-0.020	-0.006	-0.006	-0.009	-0.010
		10	-0.055	-0.053	-0.018	-0.014	-0.031
		15	-0.042	-0.038	-0.011	-0.027	-0.015
	5	5	-0.023	-0.048	0.018	0.003	-0.027
		10	-0.043	-0.027	-0.013	0.000	-0.009
		15	-0.021	-0.023	-0.007	0.002	-0.011

Note. Tau Eq. = Essentially tau-equivalent items, Congeneric = Essentially congeneric items, Var. = variance, Cov. = covariance

After removing the inadmissible solutions, the relative bias of the estimate of the mean of the level depended on the type of item (partial η^2 = 0.843), invariance (partial η^2 = 0.170), number of measurement times (partial η^2 = 0.076), and the interaction between number of measurement times and invariance (partial η^2 = 0.087). The relative bias of the estimate of the mean of the level was positive in all conditions (see Table 19). However, with essentially tau-equivalent items, the relative bias was about twice larger than the relative bias with essentially congeneric items. The overall mean of the relative biases of the estimate of the mean of the level was 0.513 with tau-equivalent items and 0.230 with essentially congeneric items. Within conditions with either essentially tau-equivalent items or essentially congeneric items, the relative bias was consistent.

The relative bias of the estimate of the mean of the shape, after removing inadmissible solutions, depended on type of item (partial η^2 = 0.942), invariance (partial η^2 = 0.641), number of measurement times (partial η^2 = 0.334), number of items (partial η^2 = 0.316), the interaction between number of times and invariance (partial η^2 = 0.321), the interaction between number of items and invariance (partial η^2 = 0.208), the interaction between number of items and number of times (partial η^2 = 0.109), and the interaction between number of items and type of items (partial η^2 = 0.083). The relative bias of the estimate of the mean of the shape was acceptable in all conditions with essentially tau-equivalent items and strict factorial invariance (see table 19). With essentially tau-equivalent items and weak factorial invariance, the relative bias was negative with three measurement times but acceptable with five measurement times. In conditions with three measurement times, five items, weak factorial invariance and essentially tau-equivalent items, the relative bias of the estimate of the mean of the shape was as large as -0.192. When the items were essentially congeneric, the estimate of the mean of the shape presented unacceptable and negative relative bias (see Table 19), regardless of the other conditions. In conditions with essentially congeneric items, the relative bias had an overall mean of -0.288.

The relative bias of the estimate of the variance of the level depended on the type of item (partial η^2 = 0.721). With essentially tau-equivalent items, the relative bias was acceptable in all conditions (see Table 19). The overall mean relative bias of the estimate of the variance of the level with essentially tau-equivalent items was 0.012. With essentially congeneric items, the relative bias was negative and consistent across all conditions (see Table 19). The overall mean relative bias with essentially congeneric items was -0.472.

The relative bias of the estimate of the variance of the shape depended on type of item (partial η^2 = 0.412) and number of measurement times (partial η^2 = 0.050). The estimate of the variance of the shape presented acceptable relative bias with essentially tau-equivalent items only when the number of measurement times was five (see Table 19). When the number of measurement times was three and the items were essentially tau-equivalent, the relative bias of the estimate of the variance of the shape was positive. In this condition, the magnitude of the relative bias was consistently around 0.17. With essentially congeneric items, the relative bias was unacceptable and negative for all conditions (see Table 19). The overall mean relative bias of the estimate of the variance of the shape with essentially congeneric items was -0.397. The magnitude of the relative biase when the items were essentially congeneric ranged from -0.110 to -0.559.

The relative bias of the estimate of the correlation between level and shape did not depend on any of the conditions included in this study. The relative bias of the estimate of the correlation between level and shape was acceptable in the conditions with strict factorial invariance and five measurement times, regardless of type and number of items, sample size or reliability (see Table 19). In the conditions with weak factorial invariance and five measurement times, the relative bias is positive and decreases as the number of items per construct increases, becoming marginally acceptable when the number of items reaches 15 (see Table 19). In conditions with three measurement times and either strict or weak factorial invariance, the relative bias was negative, and most of values of relative bias were around -0.12. Finally, with essentially congeneric items and configural invariance, the relative bias of the estimate of the correlation between level and shape was positive regardless of the other conditions.

Table 19
Mean relative bias of parameter estimates across conditions with the univariate LGM with fixed error variances, after removing the inadmissible solutions, collapsing across sample size and reliability

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. shape	Cor. level shape
Tau Eq.							
Strict	3	5	0.561	0	0.041	0.169	-0.113
		10	0.520	0	0.042	0.181	-0.122
		15	0.534	0	0.043	0.177	-0.120
	5	5	0.560	0	-0.005	-0.004	0.031
		10	0.520	0	-0.005	-0.007	0.032
		15	0.532	0	-0.007	-0.010	0.028
Weak	3	5	0.545	-0.192	0.041	0.180	-0.131
		10	0.512	-0.102	0.039	0.174	-0.113
		15	0.517	-0.065	0.037	0.172	-0.105
	5	5	0.432	-0.047	-0.047	-0.056	0.193
		10	0.442	-0.013	-0.021	-0.025	0.087
		15	0.482	-0.016	-0.013	-0.012	0.052
Congeneric							
Strict	3	5	0.260	-0.300	-0.491	-0.423	-0.109
		10	0.269	-0.251	-0.415	-0.340	-0.038
		15	0.259	-0.273	-0.451	-0.378	-0.125
	5	5	0.260	-0.300	-0.513	-0.512	0.033
		10	0.270	-0.250	-0.438	-0.441	0.029
		15	0.260	-0.273	-0.474	-0.474	0.029
Weak	3	5	0.247	-0.492	-0.493	-0.427	-0.107
		10	0.261	-0.351	-0.414	-0.333	-0.121
		15	0.244	-0.339	-0.451	-0.383	-0.115
	5	5	0.131	-0.349	-0.550	-0.559	0.329
		10	0.193	-0.264	-0.455	-0.460	0.130
		15	0.209	-0.290	-0.483	-0.481	0.071
Configural	3	5	0.244	-0.316	-0.473	-0.110	0.084
		10	0.214	-0.264	-0.491	-0.416	0.557
		15	0.222	-0.229	-0.484	-0.328	0.359
	5	5	0.194	-0.249	-0.475	-0.374	0.223
		10	0.167	-0.180	-0.484	-0.331	0.239
		15	0.233	-0.228	-0.456	-0.382	0.114

Note. Tau Eq. = Essentially tau-equivalent, Congeneric = Essentially congeneric items, var. = variance, cor. = correlation

The relative bias of the estimate of the standard errors of the mean of the level and mean of the shape did not depend on any of the conditions manipulated in this study. The relative bias was acceptable in all conditions, with the exception of the condition with essentially congeneric items, strict factorial invariance, three measurement times and 10 items, where the relative bias was marginally acceptable (see Table 20).

The relative bias of the estimate of the standard error of the variance of the level depended on the number of measurement times (partial $\eta^2 = 0.242$) and invariance (partial $\eta^2 = 0.053$). The relative bias of the estimate of the standard error of the variance of the level was acceptable in most conditions except in a few selected conditions (see table 20). In conditions with essentially congeneric items, three measurement times and configural invariance, the relative bias ranged from 0.113 to 0.277 (see Table 20).

The relative bias of the estimate of the standard error of the variance of the shape depended on the number of measurement times (partial $\eta^2 = 0.587$), invariance (partial $\eta^2 = 0.130$), sample size (partial $\eta^2 = 0.083$), number of items (partial $\eta^2 = 0.083$), the interaction between number of times and sample size (partial $\eta^2 = 0.139$), the interaction between number of items and invariance (partial $\eta^2 = 0.124$) and the interaction between number of measurement times and invariance (partial $\eta^2 = 0.120$). The estimate of the standard error of the variance of the shape was unbiased in all conditions with five measurement times, but positively biased in all conditions with three measurement times (see Table 20). In the conditions with three measurement times and either strict or weak factorial invariance, the relative bias was consistently around 0.2. In the conditions with three measurement times and configural invariance, the relative bias ranged from 0.175 to 0.686 (see Table 20).

The relative bias of the standard error of the covariance between level and shape depended only on type of item (partial η^2 = 0.240). The relative bias was acceptable in all conditions with five measurement times and positive in all conditions with three measurement times. In the conditions with three measurement times the magnitude of the relative bias ranged from 0.11 to 0.645 (see Table 20).

Table 20
Mean relative bias of the estimates of standard error across conditions with the univariate LGM with fixed error variances, after removing the inadmissible solutions, collapsing across sample size and reliability

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. Shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.002	-0.006	0.108	0.204	0.189
		10	-0.004	0.010	0.082	0.191	0.177
		15	0.000	0.004	0.080	0.209	0.192
	5	5	-0.001	-0.002	0.001	-0.019	-0.002
		10	-0.011	-0.011	-0.007	-0.008	0.004
		15	0.000	-0.011	0.022	0.004	0.003
Weak	3	5	-0.022	-0.004	0.098	0.216	0.206
		10	-0.003	-0.002	0.099	0.214	0.190
		15	-0.016	-0.003	0.083	0.213	0.181
	5	5	-0.043	-0.032	0.009	0.005	0.019
		10	-0.016	-0.034	0.013	0.004	0.000
		15	-0.006	-0.012	-0.009	0.003	0.004
Congeneric							
Strict	3	5	-0.016	0.000	0.089	0.187	0.159
		10	-0.109	-0.106	0.041	0.202	0.110
		15	-0.008	-0.010	0.085	0.195	0.180
	5	5	0.001	0.002	0.008	-0.005	0.007
		10	0.000	-0.006	-0.005	0.004	-0.006
		15	-0.011	0.009	-0.007	0.007	0.011
Weak	3	5	0.000	-0.014	0.082	0.230	0.204
		10	-0.002	-0.014	0.093	0.206	0.202
		15	0.000	-0.016	0.093	0.200	0.193
	5	5	-0.070	-0.079	0.020	0.012	0.029
		10	-0.033	-0.035	-0.011	0.007	0.014
		15	-0.011	0.000	-0.004	0.001	-0.001
Configural	3	5	-0.023	-0.006	0.113	0.175	0.182
		10	-0.030	-0.038	0.277	0.686	0.645
		15	-0.017	-0.037	0.187	0.382	0.380
	5	5	-0.022	-0.048	0.028	0.010	-0.017
		10	-0.042	-0.027	-0.001	0.008	0.004
		15	-0.020	-0.022	-0.001	0.007	-0.004

Note. Tau Eq. = Essentially tau-equivalent items, Congeneric = Essentially congeneric items, Var. = variance, Cov. = covariance

With the univariate LGM with fixed error variances, the chi-square fit statistics and the fit indices were similar regardless of whether they were calculated based on all solutions or only the admissible solutions. Therefore, the results will be presented considering only admissible solutions.

The relative bias of the chi-square statistic depended on invariance (partial η^2 = 0.335), sample size (partial η^2 = 0.285), number of measurement times (partial η^2 = 0.069), reliability (partial η^2 = 0.062), the interaction between sample size and invariance (partial η^2 = 0.235), the interaction between number of measurement times and invariance (partial η^2 = 0.192), the interaction between number of items and invariance (partial η^2 = 0.171), and the interaction between the number of items and the number of measurement times (partial η^2 = 0.078).

The chi-square statistic obtained using the univariate LGM with fixed error variances was unbiased in the conditions with strict factorial invariance, regardless of the other conditions (see Table 21). In the conditions with strict factorial invariance, the relative bias of the chi-square statistic decreased as sample size increased. However, the relative bias of the chi-square statistic was unacceptable and positive in conditions with weak or configural invariance. With weak and configural invariance, the relative bias increased consistently as sample size increased.

Table 21
Relative bias of the chi-square statistic with the univariate LGM with fixed error variances, collapsing across type of item and number of items

Reliability	Times	Sample	Invariance		
			Strict	Weak	Configural
0.70	3	100	0.018	0.394	3.637
		200	0.004	0.735	7.057
		500	0.000	1.958	18.058
		1000	0.009	3.810	36.136
	5	100	0.022	0.920	1.104
		200	0.006	1.801	2.168
		500	-0.001	4.458	5.422
		1000	0.003	8.902	10.797
0.90	3	100	0.009	0.733	6.795
		200	-0.012	1.460	13.231
		500	-0.004	3.737	33.619
		1000	0.015	7.314	67.926
	5	100	0.029	1.851	2.262
		200	0.017	3.640	4.485
		500	0.010	9.059	11.072
		1000	0.004	18.122	22.187

The CFI obtained with the univariate LGM with fixed error variances depended on invariance (Cramer's V = 0.458), type of item (Cramer's V = 0.213), number of measurement times (Cramer's V = 0.188), and number of items (Cramer's V = 0.136).

The TLI obtained with the univariate LGM with fixed error variances depended on invariance (Cramer's V = 0.526), type of item (Cramer's V = 0.251) and number of items (Cramer's V = 0.103).

The univariate LGM with fixed error variances produced similar proportions of CFI and TLI that would indicate an acceptable model fit. In conditions with strict factorial invariance, the percentage of CFI and TLI that would indicate an adequate fit of the model was at least 99.6% regardless of reliability, type of item, number of items, number of measurement times and sample size (see Table 22). With weak factorial invariance, the percentage of CFI and TLI that would indicate an acceptable model fit was below 33% if the number of times was five and the number of items was five, but it was at least 79% otherwise. With configural invariance, the percentage of CFI and TLI equal or above 0.95 was below 58% with all conditions, with the exception of the condition with congeneric items, three measurement times and five items, in which the percentage of CFI was 96.7% and the percentage of TLI was 71.8%.

The RMSEA obtained with the univariate LGM with fixed error variances depended on invariance (Cramer's V = 0.630), type of item (Cramer's V = 0.233) and number of measurement times (Cramer's V = 0.120). The percentage of RMSEA that would support the adequate fit of the model was between 82.5% and 89.8% in conditions with strict factorial invariance (see Table 22). In the conditions with weak and configural invariance, the percentage of RMSEA equal or below 0.05 varied substantially in each condition, but did not exceed 75% in any condition.

The fit of each univariate LGM with fixed error variances was also evaluated with the combination of the CFI, TLI, and RMSEA suggested by Hu and Bentler (1999). The fit of the model as measured by the combination of CFI $\geq$ 0.95, TLI $\geq$ 0.95, RMSEA $\leq$ 0.05 depended on invariance (Cramer's V = 0.630), type of item (Cramer's V = 0.234) and number of measurement times (Cramer's V = 0.120). The percentage of models whose fit indices would support the adequate fit of the model according to the Hu and Bentler's combined criterion was determined by the percentage of RMSEA equal or below 0.05, because it was lower than the percentage of CFI and TLI equal or above 0.95.

Table 22
Percentage of replications in which the fit indices would lead to model retention with the univariate LGM with fixed error variances, collapsing across sample size and reliability conditions.

Invariance	Type of Item	Times	Items	CFI	TLI	RMSEA	Combined
Strict	Tau Eq.	3	5	99.6%	96.5%	82.4%	82.4%
			10	99.6%	96.6%	82.5%	82.5%
			15	99.7%	97.1%	83.1%	83.1%
		5	5	99.8%	99.8%	89.5%	89.5%
			10	99.7%	99.7%	88.6%	88.6%
			15	99.8%	99.8%	89.5%	89.5%
	Congeneric	3	5	99.6%	96.7%	83.1%	83.1%
			10	99.6%	96.5%	82.5%	82.4%
			15	99.7%	97.1%	82.7%	82.7%
		5	5	99.9%	99.9%	89.4%	89.4%
			10	99.8%	99.8%	89.7%	89.7%
			15	99.8%	99.8%	89.8%	89.8%
Weak	Tau Eq.	3	5	98.9%	91.7%	54.8%	54.8%
			10	99.4%	95.3%	72.2%	72.2%
			15	98.6%	90.5%	49.5%	49.5%
		5	5	32.6%	32.6%	1.3%	1.3%
			10	96.4%	96.4%	14.5%	14.5%
			15	99.2%	99.2%	60.3%	60.3%
	Congeneric	3	5	98.0%	83.9%	36.9%	36.9%
			10	99.4%	93.8%	62.8%	62.8%
			15	97.7%	79.9%	32.1%	32.1%
		5	5	1.6%	1.6%	0.1%	0.1%
			10	81.7%	81.7%	4.2%	4.2%
			15	98.3%	98.3%	33.3%	33.3%
Configural	Congeneric	3	5	96.7%	71.8%	25.3%	25.3%
			10	19.8%	3.5%	1.2%	1.2%
			15	57.5%	11.3%	4.3%	4.3%
		5	5	52.5%	52.5%	2.4%	2.4%
			10	46.0%	46.0%	1.8%	1.8%
			15	50.4%	50.4%	1.9%	1.9%

Note. Criteria for model retention used to calculate the percentages:
CFI $\geq$ 0.95, TLI $\geq$ 0.95, RMSEA $\leq$ 0.05,
Combined = Criterion of acceptable fit based on the combination of CFI $\geq$ 0.95, TLI $\geq$ 0.95 and RMSEA $\leq$ 0.05 (Hu & Bentler, 1999)

Univariate LGM with the correction for attenuation

While the first three methods that have been discussed did not have any convergence problems, there were several analyses with the univariate latent growth model with the correction for attenuation that did not converge (see Table 23). The reason comes from a well-known limitation of the correction for attenuation: It sometimes produces corrected correlation coefficients that are larger than one. Consequently, some of the corrected variance/covariance matrices of composites were non-positive definite and the analyses could not be performed. The number of non-converged analyses was higher when the composite reliability was 0.7 because the lower the reliability, the larger the corrected correlation coefficient will be as compared with the uncorrected coefficient, and it is more likely that the corrected correlation will exceed one. The number of non-converged analyses was also higher in conditions with the smallest sample size (i.e. 100) and with five measurement times (see Table 23). The number of measurement times had an effect on convergence because in this study the correlation between level and shape was positive, so the inter-item correlations tended to increase with time. Consequently, the corrected correlations between the composites at later measurement times were more likely to exceed one than at earlier measurement times.

Table 23
Non-converged solutions with the univariate LGM with the correction for attenuation, collapsing across number of items, type of item, and invariance.

Sample size	Reliability	Percentage of non-convergent analyses	
		3 times	5 times
100	0.7	1.15%	35.73%
	0.9	0%	0.11%
200	0.7	0.03%	8.30%
	0.9	0%	0%
500	0.7	0%	0.35%
	0.9	0%	0%
1000	0.7	0%	0%
	0.9	0%	0%

With the univariate latent growth model with the correction for attenuation, there were also some analyses that resulted in non-positive definite estimates of the variance/covariance matrix of level and shape. The number of inadmissible solutions depended on the number of measurement times and on sample size. The percentage of these inadmissible solutions for each measurement time and sample size is displayed in table 24.

Table 24
Percentage of inadmissible solutions with the univariate latent growth model with the correction for attenuation, by sample size and measurement time, collapsing across the other conditions.

Sample size	Times		Total
	3	5	
100	30.3%	2.7%	16.5%
200	22.5%	0.8%	11.7%
500	13.7%	0%	6.8%
1000	9.4%	0%	4.7%

The relative bias of the parameter estimates and standard errors with the univariate latent growth model with the correction for attenuation was calculated both without removing inadmissible solutions and with the inadmissible solutions removed.

Because there were non-convergence problems with the univariate latent growth model with correction for attenuation, extra datasets were generate to complete 1000 converged analyses for each condition. The 1000 converged analyses for each condition were used to calculate the relative biases of the parameter estimates and standard errors without removing inadmissible solutions. Then, the inadmissible solutions were removed and extra datasets were generated to complete 1000 analyses with admissible solutions per condition. These 1000 admissible solutions were used to compute the relative bias of the parameter estimates and standard errors a second time.

Without removing the inadmissible solutions, the relative bias of the estimate of the mean of the level depended on type of item (partial η^2 = 0.810), invariance (partial η^2 = 0.198), measurement times (partial η^2 = 0.073), the interaction between number of measurement times and invariance (partial η^2 = 0.114) and the interaction between number of items and invariance (partial η^2 = 0.058). The estimate of the mean of the level was unacceptably biased in all conditions, and the relative bias was consistently positive (see Table 25). However, the relative bias of the estimate of the mean of the level in conditions with essentially tau-equivalent times was twice as large as the relative bias in conditions with essentially congeneric items. The overall mean relative bias in conditions with essentially tau-equivalent items was 0.507. In conditions with essentially congeneric items, the overall mean relative bias was 0.222.

The relative bias of the estimate of the mean of the shape depended on type of item (partial η^2 = 0.924), invariance (partial η^2 = 0.540), number of measurement times (partial η^2 = 0.301), number of items (partial η^2 = 0.239), the interaction between measurement times and invariance (partial η^2 = 0.298), the interaction between number of items and invariance (partial η^2 = 0.152), the interaction between number of items and number of times (partial η^2 = 0.107) and the interaction between number of items and type of item (partial η^2 = 0.065). The estimate of the mean of the shape was unbiased in conditions with essentially tau-equivalent items and strict factorial invariance, and also in conditions with essentially tau-equivalent items, weak factorial invariance and five measurement times. In the conditions with essentially tau-equivalent items, weak factorial invariance and three measurement times, the estimate of the mean o the shape was negatively biased, and the relative bias ranged from -0.063 with 15 items to -0.190 with 5 items. The estimate of the mean of the shape in conditions with essentially congeneric items was consistently negatively biased (see Table 25). The relative bias ranged from -0.171 to -0.490, with an overall mean of -0.286.

The relative bias of the estimate of the variance of the level depended on the type of item (partial η^2 = 0.626). All of the conditions with essentially tau-equivalent items had an acceptable relative bias of the estimate of the variance of the level. On the other hand, all of the conditions with essentially congeneric items resulted in a negatively biased estimate of the variance of the level. With essentially congeneric items, the relative bias had an overall mean of -0.487 and it was consistent across the other conditions.

The relative bias of the estimate of the variance of the shape depended on type of item (partial η^2 = 0.237). The estimate of the variance of the shape was unbiased in the conditions with essentially tau-equivalent items, but negatively biased in the conditions with essentially congeneric items (see Table 25). The overall mean relative bias was -0.013 and -0.472 with essentially tau-equivalent and essentially congeneric items, respectively. In the conditions with essentially congeneric items, configural invariance, three measurement times and 10 items, the relative bias of the estimates of the mean of the shape was as large as -0.85.

The relative bias of the estimate of the covariance between level and shape depended on the type of item (partial η^2 = 0.186) and invariance (partial η^2 = 0.055). The relative bias was acceptable in all conditions with essentially tau-equivalent items, except when there was weak factorial invariance, the number of times was five and the number of items was five (see Table 25). In the conditions with essentially tau-equivalent items, the overall mean relative bias was 0.018. With essentially congeneric items, most of the estimates of the covariance between level and shape were negatively biased (see Table 25). The only two exceptions were the conditions with essentially congeneric items, configural invariance, three measurement times, and ten items, which had positive relative bias, and the conditions with essentially congeneric items, configural invariance, three measurement times and fifteen items, which resulted in acceptable relative bias. The overall mean relative bias of the conditions with essentially congeneric items was -0.363.

Table 25
Mean relative bias of parameter estimates across conditions with the univariate LGM with the correction for attenuation, without removing the inadmissible solutions, collapsing across sample size and reliability

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. shape	Cov. level shape
Tau Eq.							
Strict	3	5	0.561	-0.001	-0.013	-0.025	0.023
		10	0.520	0.000	-0.004	-0.001	-0.004
		15	0.534	0.001	-0.009	-0.029	0.021
	5	5	0.560	0.000	0.006	0.011	-0.011
		10	0.520	0.000	0.008	0.011	-0.020
		15	0.533	0.000	0.003	0.004	-0.019
Weak	3	5	0.544	-0.190	-0.017	-0.029	0.025
		10	0.511	-0.100	-0.012	-0.020	0.015
		15	0.515	-0.063	-0.011	-0.018	0.018
	5	5	0.398	-0.031	-0.053	-0.042	0.112
		10	0.422	-0.001	-0.023	-0.016	0.047
		15	0.474	-0.010	-0.008	0.007	0.007
Congeneric							
Strict	3	5	0.260	-0.301	-0.514	-0.519	-0.506
		10	0.270	-0.250	-0.440	-0.447	-0.438
		15	0.259	-0.273	-0.478	-0.481	-0.466
	5	5	0.260	-0.300	-0.507	-0.502	-0.519
		10	0.270	-0.250	-0.433	-0.431	-0.445
		15	0.259	-0.273	-0.469	-0.466	-0.479
Weak	3	5	0.244	-0.490	-0.517	-0.525	-0.496
		10	0.259	-0.349	-0.440	-0.441	-0.433
		15	0.241	-0.336	-0.479	-0.490	-0.459
	5	5	0.091	-0.335	-0.549	-0.542	-0.426
		10	0.170	-0.252	-0.457	-0.456	-0.395
		15	0.200	-0.283	-0.479	-0.471	-0.467
Configural	3	5	0.243	-0.314	-0.499	-0.216	-0.229
		10	0.183	-0.264	-0.591	-0.850	0.248
		15	0.208	-0.227	-0.544	-0.559	-0.037
	5	5	0.177	-0.241	-0.469	-0.383	-0.307
		10	0.155	-0.171	-0.467	-0.330	-0.301
		15	0.244	-0.237	-0.443	-0.390	-0.380

Note. Tau Eq. = Essentially tau-equivalent items, Congeneric = Essentially congeneric items, var. = Variance, cov. = Covariance

The bias of the estimate of the standard error of the mean of the level depended on reliability (partial η^2 = 0.651), invariance (partial η^2 = 0.243), number of measurement times (partial η^2 = 0.125) and sample size (partial η^2 = 0.124). The estimate of the standard error of the mean of the level was negatively biased in all conditions with reliability equal to 0.7, regardless of invariance, type of item, number of items, number of measurement times or sample size (see Table 26). The overall mean relative bias in conditions with reliability equal to 0.7 was -0.277. With reliability of 0.9 and strict factorial invariance, the estimate of the standard error of the mean of the level was unbiased. With reliability of 0.9 and weak factorial invariance, the estimate of the standard error of the mean of the level was unbiased if sample size was larger than 500. With reliability of 0.9 and configural invariance, all estimates were negatively biased (see Table 26).

The bias of the estimate of the standard error of the mean of the shape depended on reliability (partial η^2 = 0.692), invariance (partial η^2 = 0.397), sample size (partial η^2 = 0.129), the interaction between number of measurement times and invariance (partial η^2 = 0.207), the interaction between number of items and number of measurement times (partial η^2 = 0.114), the interaction between number of items and invariance (partial η^2 = 0.099), the interaction between reliability and invariance (partial η^2 = 0.059), and the interaction between reliability and sample size (partial η^2 = 0.053). The estimate of the standard error of the mean of the shape was negatively biased in all conditions where reliability was 0.7 (see Table 26). The overall mean relative bias in these conditions was -0.42. In conditions with reliability of 0.9, strict factorial invariance and sample size larger than 100, the relative bias was acceptable. In the conditions with reliability of 0.9 and either weak or configural invariance, the estimate of the standard error of the mean of the shape was negatively biased (see Table 26)

The relative bias of the estimates of the standard errors of the variance of the level, variance of the shape, and covariance between level and shape were consistently unacceptable and negative in all conditions (see Table 26). The magnitude of the relative bias ranged from -0.38 and -0.57 when the reliability was 0.7. When the reliability was 0.9, the relative bias ranged from -0.13 to -0.20.

The relative bias of the estimate of the standard error of the variance of the level depended on reliability (partial η^2 = 0.668), sample size (partial η^2 = 0.070) and the interaction between number of measurement times and sample size (partial η^2 = 0.059).

The relative bias of the estimate of the standard error of the variance of the shape depended on reliability (partial η^2 = 0.740), sample size (partial η^2 = 0.108), the interaction between number of measurement times and sample size (partial η^2 = 0.085) and the interaction between number of measurement times and reliability (partial η^2 = 0.068).

The relative bias of the estimate of the standard error of the covariance between level and shape depended on reliability (partial η^2 = 0.782), sample size (partial η^2 = 0.105), the interaction between reliability and sample size (partial η^2 = 0.062) and the interaction between reliability and number of measurement times (partial η^2 = 0.052).

Table 26
Mean relative bias of the estimates of the standard errors with the univariate LGM with the correction for attenuation, without removing the inadmissible solutions, collapsing across number of measurement times, number of items, and type of item

Rel.	*Invariance*	*Sample*	*Mean level*	*Mean shape*	*Var. level*	*Var. shape*	*Cov. level shape*
0.70	Strict	100	-0.278	-0.347	-0.501	-0.531	-0.530
		200	-0.222	-0.288	-0.458	-0.495	-0.503
		500	-0.184	-0.230	-0.399	-0.428	-0.431
		1000	-0.178	-0.215	-0.390	-0.406	-0.424
	Weak	100	-0.359	-0.465	-0.476	-0.551	-0.533
		200	-0.324	-0.444	-0.447	-0.561	-0.515
		500	-0.265	-0.399	-0.388	-0.466	-0.449
		1000	-0.250	-0.355	-0.366	-0.439	-0.427
	Configural	100	-0.418	-0.512	-0.496	-0.546	-0.545
		200	-0.373	-0.483	-0.464	-0.512	-0.524
		500	-0.323	-0.429	-0.395	-0.455	-0.458
		1000	-0.297	-0.396	-0.378	-0.431	-0.439
0.90	Strict	100	-0.076	-0.102	-0.166	-0.186	-0.186
		200	-0.062	-0.083	-0.149	-0.168	-0.166
		500	-0.052	-0.075	-0.147	-0.158	-0.160
		1000	-0.054	-0.073	-0.139	-0.149	-0.151
	Weak	100	-0.114	-0.182	-0.152	-0.202	-0.183
		200	-0.105	-0.148	-0.134	-0.176	-0.172
		500	-0.093	-0.151	-0.138	-0.177	-0.167
		1000	-0.077	-0.137	-0.141	-0.168	-0.169
	Configural	100	-0.148	-0.219	-0.165	-0.206	-0.218
		200	-0.146	-0.194	-0.169	-0.190	-0.187
		500	-0.109	-0.178	-0.159	-0.180	-0.199
		1000	-0.128	-0.168	-0.137	-0.167	-0.181

Note. Rel. = Reliability of the composite, Var. = Variance,
Cov. = Covariance

After removing the inadmissible solutions and generating more datasets to obtain 1000 admissible solutions by condition, the relative bias of estimate of the mean of the level depended on type of item (partial η^2 = 0.811), invariance (partial η^2 = 0.091) and the interaction between number of measurement times and invariance (partial η^2 = 0.114). The estimate of the mean of the level was positively biased in all conditions. The overall mean of the relative bias of the mean of the level was 0.51 with essentially tau-equivalent items and 0.230 with essentially congeneric items (see Table 27).

The relative bias of the mean of the shape depended on the type of item (partial η^2 = 0.924), invariance (partial η^2 = 0.541), number of measurement times (partial η^2 = 0.294), number of items (partial η^2 = 0.241), the interaction between number of measurement times and invariance (partial η^2 = 0.295), the interaction between number of items and invariance (partial η^2 = 0.152), the interaction between number of items and number of measurement times (partial η^2 = 0.107), and the interaction between number of items and type of item (partial η^2 = 0.065). The estimate of the mean of the shape was unbiased in conditions with essentially tau-equivalent items and strict factorial invariance, or weak factorial invariance and five measurement times (see Table 27). The conditions with essentially tau-equivalent items, weak factorial invariance and three measurement times resulted in negative relative bias ranging from -0.06 to -0.190. If the items were essentially congeneric, the relative bias of the mean of the shape was negative regardless of the other conditions (see Table 27). The relative bias of the mean of the shape ranged from to -0.171 to -0.490, with an overall mean of -0.289.

The bias of the estimate of the variance of the level depended on the type of item (partial η^2 = 0.672). The relative bias was acceptable in all conditions with essentially tau-equivalent items, but was negative in conditions with essentially congeneric items (see Table 27). With essentially congenereric items, the relative bias of the estimate of the variance of the level was consistent and had an overall mean of -0.47.

The relative bias of the estimate of the variance of the shape depended on type of item (partial η^2 = 0.387). The estimate of the variance of the shape was unbiased if the items were essentially tau-equivalent and the number of measurement times was five (see Table 27). However, if the items were essentially tau-equivalent and the number of measurement times was three, the estimate of the variance of the shape was positively biased, and the magnitude of the bias was consistently around 0.18. In the conditions with essentially congeneric items, the estimate of the variance of the shape was negatively biased (see Table 27). This relative bias ranged from -0.104 to –0.53, with an overall mean of -0.397.

None of the conditions included in the study had an effect size larger than 0.05 on the relative bias of the estimate of the correlation between level and shape. The relative bias was acceptable in conditions with strict factorial invariance and five measurement times, regardless of type of item, number of items, sample size, and reliability (see Table 27). However, with three measurement times and either strict or weak factorial invariance, the relative bias was negative and of magnitude around -0.12. In conditions with five measurement times, weak factorial invariance, and either 5 or 10 items, the relative bias was unacceptable and positive, but with 15 items the relative bias was acceptable. If there was configural invariance, the estimate of the correlation between level and shape presented positive bias, ranging from 0.087 to 0.540 (see Table 27).

Table 27
Mean relative bias of parameter estimates across conditions with the univariate LGM with the correction for attenuation, after removing the inadmissible solutions, collapsing across sample size and reliability.

Type of item / Invariance	Times	Items	Mean level	Mean shape	Var. level	Var. Shape	Cor. level shape
Tau Eq.							
Strict	3	5	0.560	-0.001	0.039	0.174	-0.112
		10	0.519	0	0.046	0.185	-0.125
		15	0.534	0.001	0.042	0.181	-0.122
	5	5	0.560	0	0.011	0.015	0.009
		10	0.520	0	0.014	0.013	0.001
		15	0.533	0	0.006	0.008	0.007
Weak	3	5	0.546	-0.190	0.041	0.181	-0.125
		10	0.513	-0.100	0.039	0.182	-0.113
		15	0.517	-0.062	0.038	0.179	-0.101
	5	5	0.400	-0.031	-0.045	-0.039	0.190
		10	0.423	-0.001	-0.017	-0.009	0.087
		15	0.474	-0.011	-0.004	0.012	0.033
Congeneric							
Strict	3	5	0.261	-0.300	-0.491	-0.424	-0.112
		10	0.271	-0.250	-0.414	-0.339	-0.118
		15	0.260	-0.273	-0.452	-0.380	-0.116
	5	5	0.260	-0.300	-0.504	-0.500	0.001
		10	0.271	-0.250	-0.431	-0.430	0.001
		15	0.260	-0.273	-0.466	-0.464	0.007
Weak	3	5	0.247	-0.490	-0.491	-0.423	-0.113
		10	0.261	-0.349	-0.411	-0.329	-0.129
		15	0.244	-0.336	-0.452	-0.379	-0.110
	5	5	0.094	-0.336	-0.543	-0.538	0.278
		10	0.171	-0.253	-0.453	-0.452	0.132
		15	0.200	-0.283	-0.476	-0.469	0.048
Configural	3	5	0.245	-0.313	-0.471	-0.104	0.081
		10	0.208	-0.261	-0.489	-0.400	0.540
		15	0.220	-0.225	-0.485	-0.314	0.345
	5	5	0.177	-0.241	-0.467	-0.380	0.239
		10	0.156	-0.171	-0.463	-0.325	0.184
		15	0.244	-0.238	-0.441	-0.389	0.087

Note. Tau Eq. = Essentially tau-equivalent items, Congeneric = Essentially congeneric items, var. = variance, cor. = correlation

After removing the inadmissible solutions, the relative bias of the estimate of standard error of the mean of level depended on reliability (partial η^2 = 0.639), number of measurement times (partial η^2 = 0.204), invariance (partial η^2 = 0.191), sample size (partial η^2 = 0.125), number of items (partial η^2 = 0.067), the interaction between number of measurement times and invariance (partial η^2 = 0.123), the interaction between number of measurement times and reliability (partial η^2 = 0.111), the interaction between number of items and measurement times (partial η^2 = 0.098), the interaction between number of items and invariance (partial η^2 = 0.096) and the interaction between reliability and sample size (partial η^2 = 0.054). The estimate of the standard error of the mean of the level was negatively biased in conditions with reliability of 0.7, regardless of the other conditions (see Table 28). The relative bias tended to decrease as sample size increased, but remained unacceptable. With reliability of 0.7, the relative bias of the estimate of the standard error of the mean of the level ranged from -0.17 to -0.39, with an overall mean relative bias of -0.266. In conditions with reliability of 0.9, the relative bias was acceptable if there was strict factorial invariance, or weak factorial invariance with a sample size larger than 200 (see Table 28). In the remaining conditions, the relative bias was negative and ranged from -0.101 to -0.129.

The relative bias of the estimate of the standard error of the mean of the shape depended on reliability (partial η^2 = 0.697), invariance (partial η^2 = 0.405), sample size (partial η^2 = 0.110), the interaction between number of measurement times and invariance (partial $\eta^2 = 0.192$), the interaction between number of items and number of measurement times (partial $\eta^2 = 0.117$), the interaction between number of items and invariance (partial $\eta^2 = 0.094$), the interaction between reliability and invariance (partial $\eta^2 = 0.060$), and the interaction between reliability and sample size (partial $\eta^2 = 0.057$). The estimate of the standard error of the mean of the shape was negatively biased in all conditions with reliability of 0.7 (see Table 28). The overall mean relative bias in conditions with reliability of 0.7 was -0.363. When the reliability was 0.9 and there was strict factorial invariance, the estimate of the standard error of the mean of the shape had acceptable relative bias. However, in conditions with reliability of 0.9, if there was either weak or configural invariance, the relative bias was negative and ranged from -0.135 to -0.188.

The relative bias of the estimate of the standard error of the variance of the level depended on reliability (partial $\eta^2 = 0.620$), number of measurement times (partial $\eta^2 = 0.121$), and the interaction between number of measurement times and sample size (partial $\eta^2 = 0.117$). The estimate of the standard error of the variance of the level was negatively biased in all conditions with reliability of 0.7 (see Table 28). The overall mean relative bias in the conditions with reliability of 0.7 was -0.378. In the conditions with reliability of 0.9 and strict factorial invariance, the relative bias was negative and was consistently around -0.12. The relative bias was acceptable in the conditions with reliability of 0.9, weak factorial invariance and sample sizes of 100 and 200, as well as in the conditions with configural invariance regardless of sample size.

The relative bias of the estimate of the standard error of the variance of the shape depended on reliability (partial η^2 = 0.690), number of measurement times (partial η^2 = 0.304), invariance (partial η^2 = 0.079), number of items (partial η^2 = 0.055), the interaction between number of measurement times and sample size (partial η^2 = 0.203), the interaction between reliability and sample size (partial η^2 = 0.063), the interaction between number of measurement times and invariance (partial η^2 = 0.058), and the interaction between number of items and invariance (partial η^2 = 0.054). The estimate of the standard error of the variance of the shape was negatively biased in all conditions with reliability of 0.7 (see Table 28). The overall mean relative bias in these conditions was -0.408. With reliability of 0.9, the estimate of the standard error of the variance of the shape was unbiased if there was strict or weak factorial invariance and the sample size was bellow 500. Furthermore, the estimate was unbiased in the conditions with reliability of 0.9 and configural invariance, regardless of sample size. The conditions with reliability of 0.9, weak or strict factorial invariance and sample size of 500 or 1000 presented negative relative bias around -0.13.

The relative bias of the estimate of the standard error of the covariance between level and shape depended on reliability (partial η^2 = 0.716), number of measurement times (partial η^2 = 0.261), invariance (partial η^2 = 0.059), number of items (partial η^2 = 0.052), the interaction between number of measurement times and sample size (partial η^2 = 0.191), the interaction between reliability and sample size (partial η^2 = 0.063), and the interaction between number of items and invariance (partial $\eta^2 = 0.057$). The estimate of the standard error of the covariance between level and shape was negatively biased in all conditions where reliability was 0.7. In these conditions, the overall mean relative bias was -0407. However, the relative bias was acceptable in conditions with reliability of 0.9 and configural invariance, or conditions with weak or strict factorial invariance and sample size bellow 200. The remaining conditions presented negative bias ranging from -0.102 to -.161 (see Table 28).

Table 28
Mean relative bias of the estimates of the standard errors with the univariate LGM with the correction for attenuation, after removing the inadmissible solutions, collapsing across number of measurement times, number of items, and type of item

Rel.	Invariance	Sample	Mean level	Mean shape	Var. level	Var. shape	Cov. level shape
0.70	Strict	100	-0.270	-0.339	-0.428	-0.427	-0.437
		200	-0.222	-0.284	-0.406	-0.412	-0.424
		500	-0.184	-0.227	-0.366	-0.372	-0.383
		1000	-0.171	-0.214	-0.370	-0.373	-0.393
	Weak	100	-0.344	-0.465	-0.400	-0.452	-0.432
		200	-0.306	-0.437	-0.386	-0.475	-0.438
		500	-0.255	-0.393	-0.353	-0.409	-0.395
		1000	-0.245	-0.354	-0.343	-0.409	-0.399
	Configural	100	-0.394	-0.509	-0.408	-0.430	-0.425
		200	-0.353	-0.491	-0.391	-0.403	-0.426
		500	-0.303	-0.429	-0.330	-0.348	-0.346
		1000	-0.269	-0.400	-0.319	-0.320	-0.333
0.90	Strict	100	-0.080	-0.090	-0.112	-0.065	-0.087
		200	-0.062	-0.081	-0.115	-0.086	-0.094
		500	-0.052	-0.075	-0.132	-0.120	-0.125
		1000	-0.054	-0.075	-0.132	-0.138	-0.142
	Weak	100	-0.114	-0.183	-0.095	-0.074	-0.072
		200	-0.101	-0.151	-0.094	-0.091	-0.102
		500	-0.094	-0.148	-0.116	-0.134	-0.133
		1000	-0.075	-0.135	-0.138	-0.156	-0.161
	Configural	100	-0.129	-0.206	-0.078	-0.044	-0.051
		200	-0.124	-0.188	-0.095	-0.021	-0.026
		500	-0.097	-0.175	-0.087	-0.015	-0.048
		1000	-0.111	-0.174	-0.078	-0.009	-0.026

Note. Rel. = Reliability, Var. = Variance, Cov. = Covariance

With the univariate LGM with the correction for attenuation, the chi-square statistics and the fit indices did not change substantially with the removal of the inadmissible solutions, and therefore the results will be presented only considering admissible solutions.

The relative bias of the chi-square statistics depended on invariance (partial η^2 = 0.344), sample size (partial η^2 = 0.280), the interaction between sample size and invariance (partial η^2 = 0.232), the interaction between number of measurement times and invariance (partial η^2 = 0.174), the interaction between the number of items and invariance (partial η^2 = 0.160), and the interaction between the number of items and the number of times (partial η^2 = 0.094).

With the univariate LGM with the correction for attenuation, the relative bias of chi-square statistic was unacceptable and positive in all conditions. The relative bias of the chi-square statistic was smallest in conditions with strict factorial invariance, and largest in conditions with configural invariance. Interestingly, the relative bias decreased consistently with the increase of sample size when there was strict factorial invariance, although it remained unacceptable. On the other hand, in conditions with either weak or configural invariance, the relative bias of the chi-square statistic increased as sample size increased (see Table 29).

Table 29
Relative bias of the chi-square with the univariate LGM with the correction for attenuation, collapsing across type of item and number of items

Rel.	Times	Sample	Invariance		
			Strict	Weak	Configural
0.70	3	100	2.426	3.968	14.033
		200	2.111	4.422	23.804
		500	1.812	7.766	53.528
		1000	1.893	12.959	103.385
	5	100	7.412	10.912	11.798
		200	7.508	15.389	17.022
		500	5.886	23.715	28.246
		1000	5.496	38.503	46.321
0.90	3	100	0.514	1.716	10.555
		200	0.478	2.775	20.245
		500	0.476	6.063	50.175
		1000	0.500	11.501	99.650
	5	100	1.240	4.662	5.597
		200	1.113	7.554	9.250
		500	1.052	16.731	20.642
		1000	1.029	32.198	40.026

Rel. = Reliability

With the univariate LGM with the correction for attenuation, the CFI depended on invariance (Cramer's V = 0.507), number of measurement times (Cramer's V = 0.420), type of item (Cramer's V = 0.228), sample size (Cramer's V = 0.136), reliability (Cramer's V = 0.126), and number of items (Cramer's V = 0.100). The percentage of CFI that would support an adequate model fit ranged between 81.0% and 99.1% in conditions with strict factorial invariance. Within conditions with strict factorial invariance, if the number of measurement times was three, the percentage of CFI equal or above 0.95 was consistently around 83%. However, with strict factorial invariance and five measurement times, the percentage of CFI equal or above 0.95 was around 98%. With weak factorial invariance, the relationship between invariance and number of measurement times inverted: If the number of measurement times was three, the percentage of CFI equal or above 0.95 was around 97%, but if the number of measurement times was five, the percentage of CFI equal or above 0.95 ranged from 0% to 78.5% depending on the number of items (see Table 30). With weak factorial invariance and five measurement times, the percentage of CFI equal or above 0.95 increased as the number of items increased. With configural invariance, the percentage of CFI equal or above 0.95 was below 12% in most of the conditions (see Table 30).

The TLI depended on invariance (Cramer's V = 0.563), number of measurement times (Cramer's V = 0.267), type of item (Cramer's V = 0.263), sample size (Cramer's V = 0.169), and reliability (Cramer's V = 0.138). In conditions with strict factorial invariance and three measurement times, the percentage of TLI equal or above 0.95 was around 94%. However, in conditions with strict factorial invariance and five measurement times, the percentage of TLI equal or above 0.95 was around 82%. With weak factorial invariance and three measurement times, the percentage of TLI equal or above 0.95 ranged between 69.8% and 91.5%. With weak factorial invariance and five measurement times, the percentage of TLI equal or above 0.95 was as low as 0% in some conditions (see Table 30). With configural invariance, the percentage of TLI equal or above 0.95 was below 8% in most conditions.

The RMSEA depended on invariance (Cramer's V = 0.407), number of measurement times (Cramer's V = 0.337), type of item (Cramer's V = 0.150), and reliability (Cramer's V = 0.142). The RMSEA was more conservative in supporting the adequate fit of the model than the CFI and TLI. The percentage of RMSEA equal or below 0.05 ranged from 0% to a maximum of 67.4%. With strict factorial invariance and three measurement times, the percentage of RMSEA equal or below 0.05 was around 67%. However, with strict factorial invariance and five measurement times, the percentage of RMSEA equal or below 0.05 was around 30%. In conditions with weak and configural invariance, there was substantial variability in the percentage of RMSEA which would indicate an adequate model fit. In these conditions, the percentage of RMSEA equal or below 0.05 ranged from 0% to 52.7% (see Table 30).

The fit of the model as measured by the combination of CFI ≥ 0.95, TLI ≥ 0.95, RMSEA ≤ 0.05 depended on invariance (Cramer's V = 0.407), number of measurement times (Cramer's V = 0.337), type of item (Cramer's V = 0.150), and reliability (Cramer's V = 0.142). The percentage of analyses whose fit indices would support the adequate fit of the model based on the combined criterion was identical to the percentage of RMSEA equal or below 0.05 for all conditions (see Table 30).

Table 30
Percentage of replications in which the fit indices would lead to model retention with the univariate LGM with the correction for attenuation, collapsing across sample size and reliability conditions.

Invariance	Type of Item	Times	Items	CFI	TLI	RMSEA	Combined
Strict	Tau Eq.	3	5	98.8%	93.7%	67.5%	67.5%
			10	98.9%	94.0%	67.4%	67.4%
			15	99.1%	94.2%	67.3%	67.3%
		5	5	83.5%	83.5%	31.0%	31.0%
			10	83.5%	83.5%	30.1%	30.1%
			15	83.5%	83.5%	30.7%	30.7%
	Congeneric	3	5	98.8%	93.1%	66.0%	66.0%
			10	98.9%	93.5%	66.4%	66.4%
			15	99.0%	94.0%	67.3%	67.3%
		5	5	81.0%	81.0%	24.5%	24.5%
			10	82.7%	82.7%	28.1%	28.1%
			15	82.8%	82.8%	29.1%	29.1%
Weak	Tau Eq.	3	5	97.8%	86.7%	37.1%	37.1%
			10	98.6%	91.5%	52.7%	52.7%
			15	97.3%	84.0%	31.8%	31.8%
		5	5	1.6%	1.6%	0%	0%
			10	59.2%	59.2%	0.2%	0.2%
			15	78.5%	78.5%	4.0%	4.0%
	Congeneric	3	5	95.6%	74.6%	21.9%	21.9%
			10	98.2%	89.0%	44.5%	44.5%
			15	95.3%	69.8%	19.3%	19.3%
		5	5	0%	0%	0%	0%
			10	16.6%	16.6%	0%	0%
			15	73.8%	73.8%	0.9%	0.9%
Configural	Congeneric	3	5	93.7%	59.6%	14.9%	14.9%
			10	11.9%	2.0%	0.4%	0.4%
			15	40.7%	7.1%	2.1%	2.1%
		5	5	2.1%	2.1%	0%	0%
			10	7.9%	7.9%	0%	0%
			15	2.9%	2.9%	0%	0%

Note. Criteria for model retention used to calculate the percentages:
CFI $\geq$ 0.95, TLI $\geq$ 0.95, RMSEA $\leq$ 0.05,
Combined = Criterion of acceptable fit based on the combination of CFI $\geq$ 0.95, TLI $\geq$ 0.95 and RMSEA $\leq$ 0.05 (Hu & Bentler, 1999)

Comparison of the four methods

This study used two main criteria to evaluate the four methods of analyzing the growth of latent variables measured by multiple indicators: unbiased estimation of population parameters and their standard errors, and production of fit statistics and fit indices that would support the hypothesized model. This section will first present the differences in relative biases between analyses where the inadmissible solutions were retained and analyses where the inadmissible solutions were removed. Next, the four methods will be compared with their ability to produce unbiased estimates of the parameters and standard errors, under the many conditions considered this study. Finally, the four methods will be compared with respect to the chi-square statistic and fit indices.

Because of the fact that the maximum likelihood estimation method used sometimes produced inadmissible solutions, the relative bias of the parameter estimates was evaluated both before removing the inadmissible solutions and after removing them. In all four latent growth modeling methods evaluated, removing the inadmissible solutions did not affect substantially the distributions of the relative bias of the means of level and shape, or the distribution of the relative bias of the standard errors of the means of level and shape. Consequently, the magnitude of the relative bias of these statistics was similar both when the inadmissible solutions were retained and when they were removed. On the other hand, removing the inadmissible solutions changed the distributions of the variances and covariance of level and shape to a noticeable degree. More specifically, when the number of measurement times was three, the percentage of non-positive definite estimates of the variance/covariance matrix of level and shape was around 20%, while with five measurement times, the percentage of non-positive definite solutions was less than 1%. Therefore, removing the inadmissible solutions did not change the relative bias of the variances and covariance of level and shape when the number of measurement times was five, but changed the relative bias substantially when the number of times was three.

The effect of removing the inadmissible solutions in conditions where the number of measurement times was three with the curve-of-factors model was that the estimates of the variance of the level, variance of the shape, and covariance between level and shape, which were found to be unbiased regardless of the other conditions when the relative bias was calculated without removing inadmissible solutions, became unacceptably biased after the inadmissible solutions are removed. The same pattern was observed with the relative bias of the standard errors of the variances and covariance of level and shape.

With the univariate LGM and the univariate LGM with the fixed error variances, when inadmissible solutions are retained, the relative bias of the variances and covariance of level shape and of their standard errors are acceptable whenever the items are essentially tau-equivalent, regardless of the number of measurement times. However, after inadmissible solutions are removed, the relative biases of the variances and covariance of level and shape and of their standard errors become unacceptable with essentially tau-equivalent items whenever the number of measurement times is three. The estimates are not affected by the removal of inadmissible solutions when the number of measurement times is five.

With the univariate LGM with the correction for attenuation, when the relative biases are calculated without removing the inadmissible solutions, the estimate of the variances and covariance of level and shape is unbiased in conditions with essentially tau-equivalent items regardless of the number of measurement times. After inadmissible solutions are removed, the estimates of the variances and covariance of level and shape become unacceptably biased with essentially tau-equivalent items and three measurement times. The estimates of the standard errors of the variances and covariance of level and shape were negatively biased with the univariate LGM with the correction for attenuation, regardless of whether the inadmissible solutions were removed or not.

The four methods evaluated did not differ with respect to the percentage of inadmissible solutions. All of the methods produced around 30% of inadmissible solutions in conditions with three measurement times and around 1% of inadmissible solutions in conditions with five measurement times. However, the univariate LGM with the correction for attenuation produced non-positive definite variance/covariance matrices of the composites in addition to inadmissible solutions.

Because the entire distribution of parameter estimates should be considered when evaluating whether an estimator is unbiased, the comparison of the four methods with respect to the relative bias of parameter estimates and standard errors will only take into account the results obtained without removing inadmissible solutions.

With respect to the criterion of being able to provide unbiased parameter estimates, the curve-of-factors model, which is the only multivariate method examined in this study, performed satisfactorily under all conditions. Both the parameter estimates and standard errors were unbiased regardless of sample size, number of items, type of item, number of measurement times, invariance, and reliability.

The three univariate LGM methods performed identically with respect to estimating the means of the level and shape, because the LGM methods with fixed error variances and with correction for attenuation model did not introduce any modification to the means of the composites.

The three univariate latent growth models produced estimates of the mean of the level that were systematically different from the population mean of the level. As shown previously, this difference is due to combining items that have different factor loadings and intercepts (see equations 54 to 58), and may not affect the interpretation of the results.

With respect to the mean of the shape, biased parameter estimates would always change the interpretation of the results because they would indicate that growth was either smaller or larger that it really was in the population. The results show that all three univariate methods are only able to provide unbiased estimates of the mean of the shape when the items are essentially tau-equivalent. When the items are essentially congeneric, all of the estimates of the mean of the shape are negatively biased, regardless of the other conditions. Furthermore, with essentially tau-equivalent items, an interesting interaction was observed between number of measurement times and invariance (see Table 9): if there were only three measurement times, the relative bias of the mean of the shape was acceptable in conditions with strict factorial invariance, but not in conditions with weak factorial invariance. However, with five measurement times, the relative biases of the mean of the shape were acceptable in conditions with either strict or weak factorial invariance.

For all of the three univariate latent growth modeling methods, in conditions with three measurement times, tau-equivalent items and weak factorial invariance, the relative bias of the mean of the shape was unacceptable but approach acceptability with fifteen indicators. It can be concluded by induction that if the number of items was larger, the biases of the mean of the shape would be acceptable.

The three univariate LGM methods were able to provide unbiased estimates of the variance of the level and shape with tau-equivalent items regardless of the other conditions. However, with congeneric items, all of the univariate methods provided negatively biased estimates.

The three univariate methods produced unbiased estimates of the covariance between level and shape. The estimates of the covariance between level and shape were unbiased with essentially tau-equivalent items, but presented negative bias when the items were essentially congeneric. With essentially tau-equivalent items, the relative bias of the covariance between level and shape was only unacceptable if there was weak factorial invariance, the number of measurement times was five, and the number of items was five.

In sum, the three univariate LGM methods produced almost identical relative biases of the parameter estimates (i.e. the estimates differed by less than 0.01). Furthermore, the same pattern was observed for all parameter estimates, with the exception of the mean of the level. The pattern observed was that the estimates were unbiased with essentially tau-equivalent items and had negative biases with essentially congeneric items.

With respect to the estimation of standard errors of the means, variances and covariance between level and shape, the univariate LGM and the univariate LGM with fixed error variances produced almost identical estimates of standard errors (i.e. the estimates differed by less than 0.002). On the other hand, the univariate LGM with the correction for attenuation produced negatively biased estimates of the standard errors in conditions with reliability of 0.7 regardless of sample size, number of measurement times, number of items, type of items or invariance. In the conditions with reliability of 0.9, the univariate LGM with correction of attenuation provided an unbiased estimate of standard errors in selected conditions. Overall, the results indicate that the correction of attenuation applied to univariate LGM decreases the standard errors, which would result in an inflated type I error rate.

A comparison of mean relative bias of each parameter estimate with the four methods across conditions is presented in tables Table 31.

Table 31
Summary of relative biases of parameter estimates by type of item

Estimates	*Tau-equivalent*				*Congeneric*			
	Cufs	Ulgm	FixLgm	CorLgm	Cufs	Ulgm	FixLgm	CorLgm
Mean level	0	.513	.513	.508	0	.229	.229	.222
Mean shape	0	-.036	-.036	-.033	0	-.289	-.289	-.286
Var level	0	-.015	-.015	-.011	-.002	-.491	-.491	-.487
Var of Shape	-.001	-.023	-.023	-.012	-.002	-.473	-.473	-.472
Correlation	-.011	.024	.024	.018	-.010	-.362	-.363	-.363
SE mean level	-.003	-.010	-.010	-.155	-.006	-.021	-.021	-.202
SE mean shape	-.001	-.012	-.012	-.219	-.005	-.023	-.022	-.269
SE Var Level	-.012	-.004	-.004	-.280	-.011	-.007	-.007	-.295
SE Var shape	-.012	-.004	-.003	-.322	-.011	-.010	-.008	-.336
SE Covariance	-.010	-.004	-.003	-.316	-.014	-.009	-.009	-.335

Note. Cufs = Curve-of-factors model, Ulgm = Univariate latent growth model, FixLgm = Univariate latent growth model with fixed error variances, CorLgm = Univariate latent growth model with the correction for attenuation.

The curve-of-factors model and the three univariate LGM methods evaluated performed differently with respect to producing a chi-square statistic and fit indices that would support the fit of the model. The chi-square statistic obtained with the curve-of-factors model was only consistently unbiased with sample sizes of 500 and 1000 (see Table 32). The bias of the chi-square statistic depended on how large the model was (i.e. the number of indicators and the number of measurement times) as well as sample size. For example, in models with five items per construct and three measurement times, the relative bias of the chi-square statistic was acceptable with a sample size of 200, because the model is small. However, with the largest model (i.e. 15 items per construct and five measurement times), the chi-square statistic was only unbiased with a sample size of 1100 (see Table 6). The curve-of-factors model produced biased chi-square statistics with sample sizes of 100 and 200 in most conditions. These results agree with the finding of Jackson (2001) that for a correctly specified measurement model with 20 indicators, the chi-square statistic is positively biased with small sample sizes. The bias of the chi-square with the curve-of-factors model and small sample sizes is because with small sample sizes the parameter estimates are less stable (i.e. there is more variability in the sampling distribution of parameter estimates), resulting in larger differences between the observed and implied covariance matrices, larger discrepancy functions, and consequently larger chi-square statistics. Because the curve-of-factors model was fit to a large variance/covariance matrix of items, the differences between the implied and observed matrix accumulate, inflating the discrepancy function and the chi-square statistic. Consequently, it was observed that the relative bias of the chi-square statistic with small sample sizes increases as the number of indicators per construct and the number of measurement times increase (see Table 6), which correspond to an increase in the size of the observed variance/covariance matrix. As the sample size increases, parameter estimates become more stable, the difference between implied and observed covariance matrices is reduced, and the chi-square statistic also decreases. On the other hand, with the univariate latent growth models, the chi-square statistic was unbiased with small sample sizes because these models were fit to small variance/covariance matrices of composites.

The chi-square statistic and the fit indices obtained with the three types of univariate LGM depended most strongly on invariance (see Table 32). The relative bias of the chi-square stastistic obtained with the univariate LGM and the univariate LGM with fixed error variances was acceptable only in the conditions with strict factorial invariance. The relative bias of the chi-square obtained with the univariate LGM with correction for attenuation was not acceptable in any condition. However, many researchers do not base decisions about the fit of the model only on the value of the chi-square statistic, because it has been shown to be very sensitive to sample size if there are small model misspecifications.

Table 32
Summary of relative biases of the chi-square statistics by invariance and sample size conditions

Invariance	Sample	Cufs	Ulgm	FixLgm	Corlgm
Strict	100	0.222	0.017	0.017	2.864
	200	0.092	0.004	0.004	2.769
	500	0.033	0.001	0.001	2.303
	1000	0.016	0.010	0.010	2.232
Weak	100	0.224	0.984	0.984	5.210
	200	0.091	1.916	1.916	7.529
	500	0.033	4.809	4.809	13.576
	1000	0.017	9.553	9.553	23.785
Configural	100	0.224	3.496	3.496	10.373
	200	0.091	6.927	6.927	17.471
	500	0.034	17.260	17.260	38.229
	1000	0.017	34.608	34.607	72.962

Note. Cufs = Curve-of-factors model, Ulgm = Univariate latent growth model, FixLgm = Univariate latent growth model with fixed error variances, CorLgm = Univariate latent growth model with the correction for attenuation.

With curve-of-factors model, almost 100% of the analyses in conditions where the sample size was 500 or 1000 resulted in a CFI, TLI, and RMSEA that met their respective criterion to retain the model. However, with sample sizes of 100 and 200, the percentage of analyses whose CFI, TLI and RMSEA would lead to rejection of the model decreased as the number of items and reliability decreased (see Table 7).

The fit indices obtained with the univariate latent growth models depended most strongly on invariance (see Table 33). If there was strict factorial invariance, most of the analyses resulted in values of CFI, TLI, and RMSEA that would lead to the retention of the model, but the proportion of fit indices that would indicate inadequate fit increased substantially with weak factorial invariance. In conditions with configural invariance, most of the CFI, TLI, and RMSEA would fail to support the fit of the model.

The univariate LGM and the univariate LGM with fixed error variances performed identically with respect to the CFI, TLI and RMSEA. The method with the correction for attenuation, on the other hand, produced a smaller percentage of CFI, TLI and RMSEA that met their respective criteria for retaining the model than the other two methods. It is also interesting to notice that the RMSEA was consistently more sensitive than the other fit indices to changes in invariance.

The results have also shown that, with the univariate LGM models, if the decision to retain the model was based on the combined criterion of CFI $\geq$ 0.95, TLI $\geq$ 0.95 and RMSEA $\leq$ 0.05, as suggested by Hu and Bentler (1999), this decision would be determined solely by the value of the RMSEA, because the RMSEA was more "conservative" (i.e. produced a larger percentage of values indicating inadequate model fit) than both the CFI and TLI. However, with the curve-of-factors model, using the combined criterion would differ from using only the RMSEA to decide on the adequate fit of the model, because there were combinations of conditions where either the CFI or the TLI were more conservative than the RMSEA, so that the combined criterion would be more conservative than each of the three fit indices separately (see Table 7).

Hu and Bentler (1999) reported that with sample sizes smaller or equal 250, combination rules with either the TLI or RMSEA tend to over-reject true population models. In these cases, Hu and Bentler argue that combinations between CFI and SRMR are preferable. With the univariate latent growth models, the results of this study confirm

Hu and Bentler's (1999) findings. This study has found that with the univariate latent growth models, the RMSEA leads to model rejection more frequently than the CFI and TLI in all conditions. Furthermore, with both the curve-of-factors model and the univariate latent growth models, the TLI was found to lead to model rejection slightly more frequently than the CFI in all conditions. However, this study has found that in conditions where the curve-of-factors model is being used with sample sizes of 100 or 200, the CFI and TLI leads to model rejection more frequently than the RMSEA (see Table 33).

Table 33
Percentage of analyses whose fit indices would indicate adequate fit for each LGM method, by invariance and sample size, collapsing across other conditions

Fit index	Invariance	Sample	Cufs	Ulgm	FixLgm	CorLgm
CFI	Strict	100	39.6%	98.9%	99.0%	76.4%
		200	76.1%	99.9%	99.9%	88.0%
		500	100%	100%	100%	99.1%
		1000	100%	100%	100%	100%
	Weak	100	37.6%	80.9%	80.7%	57.1%
		200	73.7%	83.8%	83.6%	64.9%
		500	99.6%	84.9%	84.7%	73.0%
		1000	100%	85.1%	84.9%	75.7%
	Configural	100	37.4%	56.6%	56.0%	29.3%
		200	74.3%	55.5%	55.2%	28.0%
		500	99.8%	54.1%	53.2%	25.6%
		1000	100%	52.3%	50.9%	23.3%
TLI	Strict	100	38.0%	94.9%	94.8%	69.8%
		200	74.7%	98.4%	98.4%	84.7%
		500	100%	99.9%	99.9%	98.7%
		1000	100%	100%	100%	100%
	Weak	100	36.2%	73.5%	73.4%	47.6%
		200	72.4%	77.8%	77.6%	56.0%
		500	99.6%	81.1%	81.0%	66.6%
		1000	100%	83.2%	83.0%	71.6%
	Configural	100	35.8%	43.2%	42.5%	16.7%
		200	72.9%	40.3%	39.9%	14.4%
		500	99.7%	38.5%	37.7%	12.1%
		1000	100%	38.2%	36.9%	11.2%
RMSEA	Strict	100	60.9%	73.0%	72.8%	32.6%
		200	99.9%	82.3%	82.2%	38.3%
		500	100%	92.6%	92.5%	52.9%
		1000	100%	96.9%	96.9%	68.0%
	Weak	100	61.1%	42.8%	42.8%	22.9%
		200	99.9%	39.1%	38.9%	20.5%
		500	100%	31.9%	31.5%	15.6%
		1000	100%	27.9%	27.5%	11.8%
	Configural	100	61.1%	14.0%	13.8%	6.5%
		200	100%	7.4%	7.4%	3.7%
		500	100%	2.5%	2.4%	1.1%
		1000	100%	1.1%	1.0%	0.3%

Note. Cufs = Curve-of-factors mode, Ulgm = Univariate latent growth model, FixLgm = Univariate LGM with fixed error variances, CorLgm = Univariate LGM with the correction for attenuation.

Comparison of four LGM methods in ideal conditions

"Ideal" conditions were simulated where reliability was one, there was strict factorial invariance and the items were parallel. There were 24 of these ideal conditions, as mentioned in chapter III, created based on the combination of four sample sizes, three numbers of items, and two numbers of measurement times. The results of these conditions were not reported previously because they would make the ANOVAs not completely crossed. This section compares the four latent growth modeling methods with respect to these ideal conditions.

In the ideal conditions, the univariate latent growth models become identical, because the reliability is one and consequently there is no correction for attenuation and the error variance of the composite is zero.

As expected, the four methods produced unbiased parameter estimates of the mean of the shape, variance of the level, variance of the shape and covariance between level and shape in all of the ideal conditions. However, the estimates of the mean of the level were biased with the univariate latent growth models, as a consequence of summing items with different intercepts to form the composite, which re-scales the mean of the level (see Equations 54-58). A summary of the biases of parameter estimates with the ideal conditions are presented in table 34.

Table 34
Summary of relative biases in conditions with perfect reliability, parallel items and strict factorial invariance

Estimates	Cufs	Ulgm	FixLgm	CorLgm
Mean level	0	0.540	0.540	0.540
Mean shape	0	0	0	0
VAR level	0.001	0.002	0.002	0.002
VAR Shape	0.026	0.030	0.030	0.030
Correlation	0.003	0	0	0
SE mean level	-0.001	0	0	0
SE mean shape	-0.006	-0.006	-0.006	-0.006
SE VAR Level	0.008	0.011	0.011	0.011
SE VAR shape	0.040	0.043	0.043	0.043
SE Covariance	.038	0.043	0.043	0.043

Note. Cufs = Curve-of-factors mode, Ulgm = Univariate latent growth model, FixLgm = Univariate latent growth model with fixed error variances, CorLgm = Univariate latent growth model with the correction for attenuation.

In the ideal conditions, the curve-of-factors model only produces unbiased estimates of the chi-square statistic with sample sizes of 500 and 1000 (see Table 35). As discussed previously, the chi-square statistic of large models such as the curve-of-factors model requires large sample sizes in order to be unbiased. The chi-square statistic obtained with the univariate latent growth models was unbiased regardless of the sample size, because these models are small (see Table 35).

Table 35
Summary of the relative bias of the chi-square statistic in conditions with perfect reliability, parallel items and strict factorial invariance

Sample	Cufs	Ulgm	FixLgm	CorLgm
100	0.225	-0.009	0.022	0.009
200	0.092	0.024	0.013	0.010
500	0.034	0.007	-0.003	0.001
1000	0.019	0.018	0.004	0

Note. Cufs = Curve-of-factors mode, Ulgm = Univariate latent growth model, FixLgm = Univariate latent growth model with fixed error variances, CorLgm = Univariate latent growth model with the correction for attenuation.

In the ideal conditions, the CFI would support the adequate fit of the model in at least 99.9% of the analyses with any of the four methods, regardless of sample size. With the curve-of-factors model, the TLI would support an adequate fit of the model in 100% of the analyses. With the univariate latent growth models, the TLI would support the adequate fit of the model in at least 97.8% of the analyses with a sample size of 100, at least 99.6% of the analyses with the sample size of 200, and 100% of the analyses with sample size of 500 and 1000. The RMSEA for the curve-of-factors model with a sample size of 100 would support the adequate fit of the model in only 59.7% of the analyses. However, if the sample size was 200 or larger, the RMSEA would support the adequate fit of the model in 100% of the analyses. With the univariate latent growth models, the percentage of RMSEA equal or below 0.05 increased as the sample size increased (see Table 36). With a sample size of 100, only about 72.9% of the RMSEA would support the adequate fit of the model. The percentage of RMSEA equal or below 0.05 is around 96.8% in conditions with a sample size of 1000. These results agree with the findings of Hu and Bentler (1999) which show that the TLI and RMSEA lead to over-rejection of models with small sample sizes. For analyses with small sample sizes, Hu and Bentler (1999) recommend using a combination of the CFI and the standardized root mean squared residual (SRMR).

Table 36
Percentage of analyses whose fit indices would indicate adequate fit in conditions with perfect reliability, parallel items and strict factorial invariance

Fit index	Sample	Cufs	Ulgm	FixLgm	CorLgm
CFI	100	100%	100%	100%	99.9%
	200	100%	100%	100%	100%
	500	100%	100%	100%	100%
	1000	100%	100%	100%	100%
TLI	100	100%	97.9%	97.8%	97.9%
	200	100%	99.7%	99.7%	99.6%
	500	100%	100%	100%	100%
	1000	100%	100%	100%	100%
RMSEA	100	59.7%	73.3%	72.9%	73.1%
	200	100%	82.9%	82.8%	83.0%
	500	100%	93.1%	93.0%	92.9%
	1000	100%	97.0%	96.8%	96.9%
Combined	100	59.7%	73.3%	72.9%	73.1%
	200	100%	82.9%	82.8%	83.0%
	500	100%	93.1%	93.0%	92.9%
	1000	100%	97.0%	96.8%	96.9%

Note. Cufs = Curve-of-factors mode, Ulgm = Univariate latent growth model, FixLgm = Univariate latent growth model with fixed error variances, CorLgm = Univariate latent growth model with the correction for attenuation.

Chapter V: Discussion

This study was inspired by the fact that numerous researchers use univariate latent growth modeling to study the development of individuals on latent variables that are hypothesized to be measured by multiple indicators (e.g. Buist et al., 2002; Johnson, 2002; Li, F. et al., 2001; Mason, 2001; Roth et al., 2001; Willett & Keiley, 2000). The literature search conducted indicated that there has not been a methodological study evaluating the practice of using univariate latent growth models with composites of multiple items. Furthermore, there has been little research using more sophisticated multivariate latent growth models such as the curve-of-factors model (e.g. Duncan, S. C. & Duncan, 1996; Hancock et al., 2001), which would be an alternative to the use of univariate latent growth models with composites of multiples indicators. The current study compared the univariate latent growth model used with composites of multiple indicators with three other alternatives. This section will present the conclusions that can be drawn from the comparison of the four methods as well as the limitations of this study and suggestions for future research.

The results of this study have shown that the curve-of-factors model is able to produce unbiased parameter estimates with all types of items and invariance conditions, as long as the sample size is appropriate to ensure adequate fit statistics and fit indices. The reason that the curve-of-factors model performed well is that it estimates the indicator parameters (i.e. item loadings, error variances, and intercepts), which accounts for the effects of having different types of indicators (i.e. parallel, essentially tau-equivalent or essentially congeneric) and different invariance conditions. This study has shown that without a sample size of at least five hundred, however, the fit statistics and fit indices of the curve-of-factors model tend not to be adequate, although the parameter estimates may be unbiased. On the other hand, if the number of indicators is small (i.e. five or fewer) the curve of factors model may be used with sample sizes smaller than five hundred without compromising the adequacy of fit statistics and fit indices. The sample size requirement of the curve-of-factors model may represent a limitation for applied researchers who collect their own data, because sometimes it can be difficult to obtain large samples. Furthermore, educational and psychological scales frequently contain more than five items, which would increase the minimum sample size requirement.

The three univariate methods compared in this study performed similarly, with a few differences. The results of this study indicate that using the univariate latent growth model with composites of multiple indicators, which is the practice most commonly found in the applied literature, produces biased parameter estimates if the indicators are essentially congeneric. The univariate latent growth models with fixed error variances and with the correction for attenuation also produce biased parameter estimates with essentially congeneric items. If the indicators are essentially tau-equivalent, all three univariate LGM methods produced unbiased estimates of the mean of the shape, variance of the level, variance of the shape, and covariance between level and shape. The three univariate LGM methods produce biased estimates of the mean of the level with either essentially tau-equivalent or essentially congeneric items, but this bias will not affect the interpretation of the results for a single group of individuals if the researcher does not interpret the magnitude of the mean of the shape as a percentage of the mean of the level.

The univariate latent growth model and the univariate latent growth model with fixed error variances were found to produce unbiased estimates of standard errors in all conditions (see Tables 10 and 18), while the univariate latent growth model with the correction for attenuation produces negatively biased estimates of standard errors (see Table 26).

This study has shown that the univariate latent growth model with fixed error variances produces the same parameter estimates, standard errors, chi-square statistic and fit indices as the univariate latent growth model. The reason is presented below in equations 59 to 64. Equation 59 is the general expression for the variance of the composite as a function of the variances of the level, shape, and error, and the covariance between level and shape, with the univariate latent growth model presented in Figure 5.

$$\sigma^2_{C_j} = \sigma^2_{\zeta_\alpha} + \lambda_j^2 \sigma^2_{\zeta_\beta} + 2\lambda_j \sigma_{\zeta_\alpha \zeta_\beta} + \varepsilon_j \qquad \textbf{(59)}$$

Where $\sigma^2_{C_j}$ is the variance of the composite C_j, $\sigma^2_{\zeta_\alpha}$ is the variance of the level, λ_j is the factor loading of the composite j on the shape factor, $\sigma^2_{\zeta_\beta}$ is the variance of the shape, ε_j is the error variance, and $\sigma_{\zeta_\alpha \zeta_\beta}$ is the covariance between level and shape.

Equation 60 is the general expression for the variance of the composite as a function of the variance of the latent construct and the reliability of the composite, with the univariate latent growth model with fixed error variances presented in Figure 6.

$$\sigma^2_{C_j} = \sigma^2_{\eta_j} + (1 - \rho_{c_j c_j'}) \sigma^2_{C_j} \qquad \textbf{(60)}$$

Where $\sigma^2_{\eta_j}$ is the variance of the latent construct, $\rho_{c_j c_j'}$ is the reliability of the construct, and $(1-\rho_{c_j c_j^{'}})\sigma^2_{C_j}$ is the estimated value of the error variance of the composite, which is fixed in the model.

Equation 61 is the general expression for the variance of the latent factor as a function of the variances of the level, shape and disturbance, with the univariate latent growth model with fixed error variances presented in Figure 6.

$$\sigma^2_{\eta_j} = \sigma^2_{\zeta_\alpha} + \lambda^2_j \sigma^2_{\zeta_\beta} + 2\lambda_j \sigma_{\zeta_\alpha \zeta_\beta} + \zeta_j \tag{61}$$

Equation 62 combines equations 60 and 61 to obtain the expression for the variance of the composite as a function of the variances of the level, shape and disturbance, with the univariate latent growth model with fixed error variances.

$$\sigma^2_{C_j} = \sigma^2_{\zeta_\alpha} + \lambda^2_j \sigma^2_{\zeta_\beta} + 2\lambda_j \sigma_{\zeta_\alpha \zeta_\beta} + \zeta_j + (1-\rho_{c_j c_j^{'}})\sigma^2_{C_j} \tag{62}$$

Joining equations 59 and 62 and solving for the error variance of the composite yields:

$$\varepsilon_j = \zeta_j + (1-\rho_{c_j c_j^{'}})\sigma^2_{C_j} \tag{63}$$

Equation 63 shows that in the univariate latent growth model with fixed error variances, once the error variance of the composite is fixed at $(1-\rho_{c_j c_j^{'}})\sigma^2_{C_j}$, the disturbance ζ_j is estimated so that its sum with the fixed term equals ε_j. Since ε_j and $\zeta_j + (1-\rho_{c_j c_j^{'}})\sigma^2_{C_j}$ are equal, formulas 59 and 62 become identical and produce the same estimates of the variance of the level and variance of the shape for both the univariate latent growth model and the univariate latent growth model with fixed error variances. Furthermore, for both models (see Figures 5 and 6), the general expression for the covariance between two composites *j* and *k* is:

$$\sigma_{C_j C_k} = \sigma^2_{\zeta_\alpha} + \lambda_j \lambda_k \sigma^2_{\zeta_\beta} + \lambda_j \sigma_{\zeta_\alpha \zeta_\beta} + \lambda_k \sigma_{\zeta_\alpha \zeta_\beta} \qquad \textbf{(64)}$$

Where $\sigma^2_{\zeta_\alpha}$ is the variance of the level, λ_j is the factor loading of the composite *j* on the shape factor, λ_k is the factor loading of the composite *k* on the shape factor, $\sigma^2_{\zeta_\beta}$ is the variance of the shape, and $\sigma_{\zeta_\alpha \zeta_\beta}$ is the covariance between level and shape.

The covariance between level and shape appears in the equation for the covariances between composites (see equation 64) as well as the equation for the variances of the composites (see equations 59 and 62). Because the equation for the covariances between composites is the same for both the univariate LGM and the univariate LGM with fixed error variances, and the equations for the variances of the composites for these two models (i.e. equations 59 and 62) are equivalent, the same estimate of the covariance between level and shape is obtained for both models.

In conclusion, estimating and fixing the error variances of the composites in a latent growth model will not change the estimates of the variance of the level, variance of the shape, and covariance between level and shape. However, if any factor loadings of the shape are freely estimated, the unstandardized loadings will be the same in both models, but the standardized loadings will differ. In the univariate LGM, the standardized loading of the shape is obtained by multiplying the unstandardized loading by the ratio of the standard deviation of the shape and the model implied standard deviation of the composite (i.e. $\lambda_j \sigma_{\zeta_\beta} / \sigma_{C_j}$). In the univariate LGM with fixed error variances, the standardized loading of the shape is obtained by multiplying the unstandardized loading by the ratio between the standard error of the shape and the standard deviation of the latent factor (i.e. $\lambda_j \sigma_{\zeta_\beta} / \sigma_{\eta_j}$). Because the standard deviation of the composite, σ_{C_j} , is larger than the standard deviation of the latent factor, σ_{η_j} , (unless there is perfect reliability, in which case they will be identical) the standardized loading of the shape obtained with univariate LGM is smaller than the standardized loading obtained with the univariate LGM with fixed error variances.

Some of the analyzes performed in this study with the three univariate methods, as well as with the curve-of-factors model, resulted in inadmissible solutions, which are solutions where the variance/covariance matrix of the level and shape is non-positive definite. A positive definite correlation, variance/covariance, or moment matrix is a *p x p* matrix that has all of its *p* eigenvalues greater than zero (Wothke, 1993). There are several possible causes for the matrix of variance/covariance estimates to be non-positive definite. First, the probability of having a non-positive definite solution is higher if the sample size and/or the number of indicators is small (Boomsma, 1985). Furthermore, outliers, non-normality in the data, too many parameters, and empirical under-identification can lead to inadmissible solutions (Wothke, 1993). In this study, the number of inadmissible solutions was higher when sample sizes were small, which agrees with the findings of Boomsma (1985).

Because the data were randomly generated, it is likely that empirical under-identification happened in some analyses. Empirical under-identification may happen when some of the correlations between observed variables are zero, but the model hypothesizes a non-zero parameter which depends on these correlations in order to be estimated. In this situation, the non-zero parameter specified by the model cannot be uniquely estimated from the zero observed correlations. Wothke (1993) provides an example of when empirical under-identification would appear: A model contains four indicators, and each pair is hypothesized to measure a different latent construct. The two latent constructs in the model are hypothesized to be correlated. Without any further constraints, this model has nine parameters to be estimated and one degree of freedom. However, if the observed correlations between the variables that are measuring different latent constructs are zero, the correlation between the two latent constructs cannot be estimated. With randomly simulated data, some zero correlations that would lead to empirical under-identification are likely to appear.

Because the four methods were applied to the same datasets, the percentage of inadmissible solutions was similar with all four methods. Finally, it was observed that the datasets with three measurement times produced more inadmissible solutions than the datasets with five measurement times. The larger number of inadmissible solutions with models with three measurement times may be because these models have just one degree of freedom, so the number of parameters to be estimated is very close to the number of sample statistics used to estimate them. Furthermore, the only degree of freedom of these models comes from the means part of the model, and the variance/covariance part of the model is just-identified, which would increase the chance that the simulated data would produce empirical under-indentification.

In all of the univariate LGM methods investigated, the factor that most strongly influenced the relative bias of the estimates was clearly the type of item. All of the three univariate methods tested only produced unbiased parameter estimates of the mean of the shape, variances of the level and shape, and correlation between level and shape when the items were essentially tau-equivalent. This requirement creates problems for applied researchers who want to choose the best method for their data because it is not always possible to identify in advance whether a certain set of items is essentially tau-equivalent. Furthermore, because essential tau-equivalence requires that the items have identical loadings, it may be difficult to find items which have this characteristic. It is very desirable that an analysis method is able to provide unbiased parameter estimates with essentially congeneric items, which are the most general type (Jöreskog, 1971; Millsap & Everson, 1991).

Overall, the curve-of-factors model was found to be the most adequate method to study the growth of a latent construct measured by multiple indicators. However, because the curve-of-factors model was found not to provide adequate fit statistics with small sample sizes unless the number of items is also small (e.g., 5 items per construct), a researcher who would like to study the growth of latent constructs measured by multiple indicators would have to collect enough observations (e.g., 500) to guarantee that the fit statistics and fit indices would not fail to identify a well-fitting model.

In order to identify guidelines to determine an adequate sample size for using the curve-of-factors model, the percentage of curve-of-factor models that would have an acceptable fit based on the combined criterion of CFI ≥ 0.95, TLI ≥ 0.95 and RMSEA ≤ 0.05 was calculated for each combination of number of items, number of measurement times, and sample size. For each of these combinations, the ratio of observations per parameter estimated (i.e. N:q ratio) was also calculated. The results are displayed in Table 37.

Table 37
Percentage of models that would have acceptable fit based on the combined criterion

Times	Items	Sample	Parameters	N:q ratio	Combined
3	5	100	47	2.13	82.0%
		200	47	4.26	99.1%
		500	47	10.64	100%
		1000	47	21.28	100%
	10	100	92	1.09	52.0%
		200	92	2.17	92.2%
		500	92	5.43	100%
		1000	92	10.87	100%
	15	100	137	0.73	6.6%
		200	137	1.46	66.6%
		500	137	3.65	99.7%
		1000	137	7.30	100%
5	5	100	75	1.33	77.2%
		200	75	2.67	99.7%
		500	75	6.67	100%
		1000	75	13.33	100%
	10	100	150	0.67	4.1%
		200	150	1.33	72.4%
		500	150	3.33	100%
		1000	150	6.67	100%
	15	100	225	0.44	0%
		200	225	0.89	24.8%
		500	225	2.22	98.9%
		1000	225	4.44	100%

Note. Parameters = Number of parameters estimated,
N:q ratio = Ratio of observations per parameters estimated,
Combined = Percentage of analyses that would find support for the model based on the combined criterion of CFI $\geq$ 0.95, TLI $\geq$ 0.95, and RMSEA $\leq$ 0.05

Table 37 indicates that the percentage of analyses which would result in acceptable fit based on the combined criterion can be as high as 98.9% with a N:q ratio as low as 2.22 when the sample size is 500. It is possible to conclude that the N:q ratio is not the main factor determining the percentage of fit indices which indicate acceptable models. If the N:q ratio was a strong determinant of the percentage of fit indices supporting adequate model fit, the models with small N:q ratio would be consistently associated with small percentages of fit indices indicating acceptable models. Furthermore, because the conditions with five measurement times, 10 items, sample size of 200, and N:q ratio of 2.67 resulted in a percentage of acceptable models of 99.7%, there is evidence that the percentage of acceptable models does not depend exclusively on the sample size. The table above does not allow the elaboration of a rule-of-thumb with respect to sample size requirements for the curve-of-factors model, and therefore it indicates that further research is needed to produce guidelines with respect to sample sizes for fitting the curve-of-factors model.

The univariate latent growth model with the correction for attenuation produced similar estimates of means, variances, and covariances to those of the other two univariate methods, but it severely underestimated the standard errors of the estimates. This study has shown that applying the correction for attenuation before fitting the model results in standard errors that are too small, which in turn inflates the type I error rate of hypothesis tests about the estimated parameters. These results lead to the conclusion that the correction for attenuation should not be used with univariate latent growth modeling. This conclusion implies that the recommendation of Fan (2003b) of using the correction for attenuation to estimate the true correlation between latent variables should not be applied to latent growth modeling. In the simulation study executed by Fan, he compared the performance of SEM and the correction for attenuation as separate methods to estimate the true correlation between latent constructs. He reports that both methods are able to adequately estimate the correlation between the latent constructs, but recommends that the correction for attenuation should be preferred because it is simpler to implement than SEM. Because in Fan's study he used SEM and the correction for attenuation to accomplish the same goal (i.e. estimate the correlation between latent factors), Fan does not suggest combining the correction for attenuation with SEM. This study attempted this unique idea under the hypothesis that combining the correction of attenuation with univariate LGM could improve the results of the later method. However, this study has found that univariate LGM is already able to produce unbiased estimates of standard errors in all conditions, so correcting for attenuation overdoes the work of the univariate latent growth model by reducing the standard errors too much. Because this study has found that under-estimation of standard errors happens when LGM and the correction for attenuation are combined, it is possible to conclude that it is not advisable to combine these two methods. An additional problem of the correction for attenuation found in this study is that it produces non-positive variance/covariance matrices of composites. These non-positive definite variance/covariance matrices of composites occurred because the adjusted correlation exceeded one.

None of the univariate methods examined produced unbiased parameter estimates with the most general type of items: essentially congeneric items. Because it is not possible to know in advance whether the items fit the more restricted conditions of essentially tau-equivalence, a method that provides adequate results under the most general condition should be preferred. Consequently, the recommendation that can be made based on the results of this study is that the curve-of-factors model should be used whenever the sample size is large enough to ensure adequate fit statistics and fit indices.

Limitations and suggestions for future research

The data simulation and analyses accomplished in this study have provided some important guidelines with respect to which method should be preferred to model the growth of latent constructs measured by multiple indicators. However, because in a simulation study only a certain set of conditions can be included, there are other interesting conditions that were not examined.

For example, this study only examined continuous indicators with multivariate normal distributions and no missing data. However, neither the curve-of-factors model nor the univariate latent growth models are limited to continuous and multivariate distributed data with no missing data. The choice of using continuous multivariate normally distributed indicators in this study allowed the used of maximum likelihood estimation, which is a widely known estimation method. A different estimation method which can handle categorical indicators (i.e. the WLSMV) or an estimation method which can handle missing data (i.e. the FIML) could be used to fit the curve-of-factors model or the univariate latent-growth model. It is expected that the conclusions of this study with respect to the effects of type of item and invariance on the bias of parameter estimates, standard errors and chi-square statistics, and on the fit indices would generalize for models for categorical indicators using WLSMV and models for datasets with missing data using FIML because the same relationships between the item parameters within a factor (i.e. parallel, tau-equivalent or congeneric) and the same invariance relationships (i.e. strict, weak and configural) would be observed. However, the results with respect to the effect of sample size on the relative bias of parameter estimates, standard errors and chi-square statistic, and on the fit indices would probably differ with WLSMV and FIML estimation because these estimation methods may require different sample sizes for fitting the same model.

Another limitation of this study is that only linear growth was simulated and the measurement errors of the indicators were uncorrelated. Both curvilinear growth and correlated measurement errors lead to a vast number of modeling possibilities, and including them in this study would have extended the simulation beyond manageable proportions. These are important topics which need to be addressed in future research employing the curve-of-factors model.

This study provided some information about the sample size requirements of the curve-of-factors model. However, this study did not manipulate the size of the parameters and calculate the power obtained with each method and a given sample size, number of items, and number of measurement times, because generating more than one set of parameters at least doubles the size of the simulation. A future study could restrict some of the conditions of this study but generate different sets of parameters to look more specifically into power for estimating the univariate LGM model and the curve-of-factors model.

This study also did not test slightly misspecified models, such as ones where the residual errors correlate with each other but the model presents them as uncorrelated. A study simulating misspecified models could indicate how sensitive the univariate latent growth model and the curve-of-factors model are to misspecifications.

General conclusion

This study indicates that the common practice of combining items of educational and psychological scales into composites and using univariate latent growth modeling to evaluate the level and shape of growth is only appropriate under certain conditions. Before deciding which latent growth model to use, an applied researcher should consider whether the information available about the psychometric properties of the items allows the assumption that they are essentially tau-equivalent. If the items can be assumed essentially tau-equivalent, the univariate latent growth model may be employed. On the other hand, if the items are probably congeneric and a sample size around 500 is available (if the number of indicators is small, sample sizes smaller than 500 may suffice) the researcher may use the curve-of-factors model, which does not require the very restrictive assumption of essential tau-equivalence.

Appendix A: Summary of study design

Methods of latent growth modeling:

1. LGM of item means
2. LGM of item means, with correction for attenuation
3. LGM of item means, with fixed error variances.
4. The curve-of-factors model

Invariance:

1. Strict factorial invariance
2. Weak factorial invariance
3. Configural invariance

Number of items per factor:

1. $k = 5$
2. $k = 10$
3. $k = 15$

Type of items:

1. Essentially tau-equivalent
2. Essentially congeneric

Reliability ($\rho_{C_j C_j^{'}}$):

1. $\rho_{C_j C_j^{'}} = 0.7$
2. $\rho_{C_j C_j^{'}} = 0.9$

Number of measurement times (*j*):

1. $j = 3$
2. $j = 5$

Sample size (N):

1. N = 100
2. N = 200
3. N = 500
4. N = 1000

The total number of conditions is 240.

Appendix B: Summary of ideal conditions

Invariance:

Strict factorial invariance

Type of items:

Parallel

Reliability ($\rho_{C_j C_j^{'}}$:

$\rho_{C_j C_j^{'}} = 1$

Number of items per factor:

1. k = 5

2. k = 10

3. k = 15

Number of measurement times (j):

1. j = 3
2. j = 5

Sample size (N):

1. N = 100
2. N = 200
3. N = 500
4. N = 1000

The total number of ideal conditions is 24.

Appendix C: Population parameters used in the simulation

Mean of the level = 1
Mean of the shape = 1
Variance of the level = 0.5
Variance of the shape = 0.1
Correlation between level and shape = 0.4
Covariance between level and shape = 0.089

The loadings presented below are for essentially congeneric items. The conditions with essentially tau-equivalent or parallel items had loadings equal to 1 for all items in all measurement times.

Table 38
Population factor loadings for conditions with configural invariance

Items	Time 1	Time 2	Time 3	Time 4	Time 5
Item 1	1	1	1	1	1
Item 2	0.6	0.6	0.8	0.6	0.7
Item 3	0.7	0.8	0.6	1	0.5
Item 4	0.5	0.6	0.9	0.6	0.9
Item 5	0.7	0.9	0.8	0.7	0.8
Item 6	0.7	0.7	0.9	0.7	0.9
Item 7	0.9	0.5	0.6	1	0.9
Item 8	0.6	0.8	0.9	0.8	0.6
Item 9	0.8	0.5	1	0.9	0.8
Item 10	1	0.9	0.7	1	0.8
Item 11	0.7	0.9	0.6	0.8	0.6
Item 12	0.9	0.5	0.9	0.5	1
Item 13	0.7	1	0.7	0.9	0.7
Item 14	0.6	0.7	0.8	0.8	0.6
Item 15	0.5	0.8	0.9	0.9	0.5

Table 39
Population factor loadings for conditions with weak or strict invariance

Items	Time 1	Time 2	Time 3	Time 4	Time 5
Item 1	1	1	1	1	1
Item 2	0.6	0.6	0.6	0.6	0.6
Item 3	0.7	0.7	0.7	0.7	0.7
Item 4	0.5	0.5	0.5	0.5	0.5
Item 5	0.7	0.7	0.7	0.7	0.7
Item 6	0.7	0.7	0.7	0.7	0.7
Item 7	0.9	0.9	0.9	0.9	0.9
Item 8	0.6	0.6	0.6	0.6	0.6
Item 9	0.8	0.8	0.8	0.8	0.8
Item 10	1	1	1	1	1
Item 11	0.7	0.7	0.7	0.7	0.7
Item 12	0.9	0.9	0.9	0.9	0.9
Item 13	0.7	0.7	0.7	0.7	0.7
Item 14	0.6	0.6	0.6	0.6	0.6
Item 15	0.5	0.5	0.5	0.5	0.5

Table 40
Population intercepts for conditions with configural or weak invariance

Items	Time 1	Time 2	Time 3	Time 4	Time 5
Item 1	0	0	0	0	0
Item 2	0.3	0.5	0.1	0.2	1
Item 3	0.8	0.2	0.4	0.2	0.8
Item 4	0.9	0.7	0	0.4	0.4
Item 5	0.8	0.2	0.4	0.4	0.2
Item 6	0.3	0.1	0.5	0.4	0.6
Item 7	0.4	0.3	0.7	0.5	0.9
Item 8	0.5	0.2	0.4	0.4	0.5
Item 9	0.8	0.7	0.6	0.9	0.8
Item 10	0.4	1	0.1	0.3	0.1
Item 11	0.3	0.6	0.8	1	0.2
Item 12	0.7	0.6	0.1	0.6	0.4
Item 13	0.5	0.2	0.9	0.8	0.7
Item 14	0.8	0.2	0.1	0.6	0
Item 15	0.5	0.7	1	0.1	0.2

Table 41
Population intercepts for conditions with strict invariance

Items	Time 1	Time 2	Time 3	Time 4	Time 5
Item 1	0	0	0	0	0
Item 2	0.3	0.3	0.3	0.3	0.3
Item 3	0.8	0.8	0.8	0.8	0.8
Item 4	0.9	0.9	0.9	0.9	0.9
Item 5	0.8	0.8	0.8	0.8	0.8
Item 6	0.3	0.3	0.3	0.3	0.3
Item 7	0.4	0.4	0.4	0.4	0.4
Item 8	0.5	0.5	0.5	0.5	0.5
Item 9	0.8	0.8	0.8	0.8	0.8
Item 10	0.4	0.4	0.4	0.4	0.4
Item 11	0.3	0.3	0.3	0.3	0.3
Item 12	0.7	0.7	0.7	0.7	0.7
Item 13	0.5	0.5	0.5	0.5	0.5
Item 14	0.8	0.8	0.8	0.8	0.8
Item 15	0.5	0.5	0.5	0.5	0.5

The population error variances of the items were randomly chosen given the desired composite reliability (i.e. 0.7 or 0.9) and the sum of the sum of the squared factor loadings (see equation 49) as detailed in chapter III.

References

Acock, A., & Stavig, G. R. (1979). A measure of association for nonparametric statistics. *Social Forces, 57*(4), 1381-1386.

Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: John Wiley & Sons.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423.

Argyrous, G. (2000). *Statistics for social and health research*. Thousand Oaks, CA: Sage.

Baer, J., & Schmitz, M. F. (2000). Latent growth curve modeling with a cohort sequential design. *Social Work Research, 24*(4), 243.

Bandalos, D. L., Finney, S. J., & Geske, J. A. (2003). A Model of Statistics Performance Based on Achievement Goal Theory. *Journal of Educational Psychology, 95*(3), 604-616.

Bandalos, D. L., Yates, K., & Thorndike-Christ, T. (1995). Effects of Math Self-Concept, Perceived Self-Efficacy, and Attributions for Failure and Success on Test Anxiety. *Journal of Educational Psychology, 87*(4), 611-623.

Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62*(4), 603.

Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*, *1*(3), 343-357.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement, 62*(4), 570.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*(1), 30-52.

Bollen, K. A. (1989). *Structural equations with latent variables*: John Wiley & Sons.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107*(2), 256-259.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in Lisrel maximum likelihood estimation. *Psychometrika*, *50*(2), 229-242.

Brown, J. D. (1999). *The relationship between the dimensions of self-concept and the dimensions of parental involvement: Racial/Ethnic differences and the impact on academic achievement.* Howard University, Washington, DC.

Buist, K. L., Dekovi Cacute, M., Meeus, W., & van Aken, M. A. G. (2002). Developmental patterns in adolescent attachment to mother, father and sibling. *Journal of Youth and Adolescence, 31*(3), 167.

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics, 2*(3), 177.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Calender, J. C., & Osburn, H. G. (1977). A method for maximizing and cross-validating split-half reliability coefficients. *Educational and Psychological Measurement, 37*, 819-826.

Chan, D., Ramey, S., Ramey, C., & Schmitt, N. (2000). Modeling intraindividual changes in children's social skills at home and at school: A multivariate latent growth approach to understanding between-settings differences in children's social skill development. *Multivariate Behavioral Research, 35*(3), 365-396.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1-27.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*, 107-112.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Crawford, T. N., Cohen, P., Johnson, J. G., Sneed, J. R., & Brook, J. S. (2004). The course and psychosocial correlates of personality disorder symptoms in adolescence: Erikson's developmental theory revisited. *Journal of Youth and Adolescence, 33*(5), 373-387.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. *New methods for the analysis of change.*, 107-135.

DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods, 3*(4), 412-423.

DuBois, D. L., Burk-Braxton, C., Swenson, L. P., Tevendale, H. D., Lockerd, E. M., & Moran, B. L. (2002). Getting by with a little help from self and others: Self-esteem and social support as resources during early adolescence. *Developmental Psychology, 38*(5), 822.

Duncan, S. C., & Duncan, T. E. (1996). A multivariate latent growth curve analysis of adolescent substance use. *Structural Equation Modeling, 3*(4), 323-347.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah: Lawrence Erlbaum Associates.

Fan, X. (2001). Parental involvement and students' academic achievement: A growth modeling analysis. *Journal of Experimental Education, 70*(1), 27.

Fan, X. (2003a). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Structural Equation Modeling, 10*(3), 380-400.

Fan, X. (2003b). Two approaches for correcting correlation attenuation caused by measurement error: Implications for research practice. *Educational and Psychological Measurement, 63*(6), 915-930.

Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003, 2003/04//). *The effect of sample size on latent growth models.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Hancock, G. R., & Freeman, M. J. (2001). Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educational and Psychological Measurement, 61*(5), 741.

Hancock, G. R., Kuo, W., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling, 8*(3), 470-489.

Hancock, G. R., Stapleton, L. M., & Arnold-Berkovits, I. (2004). *The tenuousness of invariance tests within multisample covariance structure models*.Unpublished manuscript, College Park, MD.

Hays, W. L. (1973). *Statistics for the social sciences*. New York, NY: Holt, Rinehard and Winston.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging, 18*(4), 639-657.

Hill, A. T. (2003). *A panel study of the effects of teacher education, class size, and time-on-task on student achievement: Evidence from NELS:88.* Unpublished Doctoral Dissertation, University of Delaware, Newark, DE.

Holt, J. K., & Stuckey, A. (2004). *Growth transition modeling: A decision matrix for analyzing multiphase longitudinal data.* Paper presented at the Annual Meeting of the American Education Research Association, San Diego, CA.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research, 26*(3), 329-367.

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

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1*(4), 179-188.

Hser, Y. (2001). Analytic approaches for assessing long-term treatment effects. *Evaluation Review, 25*(2), 233.

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus... *Structural Equation Modeling, 6*(1), 1.

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A monte carlo investigation. *Structural Equation Modeling, 8*(2), 205-223.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10*(1), 128-141.

Johnson, M. K. (2002). Social origin, adolescent experiences, and work value trajectories during the transition to adulthood. *Social Forces, 80*(4), 1307.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*(2), 109-133.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Kaplan, D. (1989). Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance. *Educational and Psychological Measurement, 49*(3), 579-586.

Kaplan, D. (1995). Statistical power in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 100-117). Thousand Oaks, CA: Sage Publications.

Kieffer, K. M., & Reese, R. J. (2002). A reliability generalization study of the geriatric depression scale. *Educational and Psychological Measurement, 62*(6), 969.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Komaroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement, 21*(4), 337-348.

Lawrence, F. R., & Hancock, G. R. (1998). Assessing change over time using latent growth modeling. *Measurement and Evaluation in Counseling and Development, 30*(4), 211-225.

Lei, P. W. (2002). Power estimation methods in structural equation modeling. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(7), 3415.

Li, F., Duncan, T. E., Duncan, S. C., McAuley, E., Chaumeton, N. R., & Harmer, P. (2001). Enhancing the psychological well-being of elderly individuals through tai chi exercise: A latent growth curve analysis. *Structural Equation Modeling, 8*(1), 53-83.

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods, 1*(1), 98-107.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53-76.

Little, T. D., Schnabel, K. U., & Baumert, J. (Eds.). (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lomax, R. G. (1986). The effect of measurement error in structural equation modeling. *Journal of Experimental Education, 54*(3), 157-162.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32*(3), 215–253.

Mason, W. A. (2001). Self-esteem and delinquency revisited (again): A test of Kaplan's self-derogation theory of delinquency using latent growth curve modeling. *Journal of Youth and Adolescence, 30*(1), 83.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Pacific Grove, CA: Brooks/Cole Publishing Company.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. *Handbook of multivariate experimental psychology (2nd ed.).* 561-614.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*(1), 110-133.

McArdle, J. J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.

McWhirter, E. H., Hackett, G., & Bandalos, D. L. (1998). A Causal Model of the Educational Plans and Career Expectations of Mexican American High School Girls. *Journal of Counseling Psychology, 45*(2), 166-181.

Meeus, W., Branje, S., & Overbeek, G. (2004). Parents and partners in crime: A six-year longitudinal study on changes in supportive relationships and deliquency in adolescence and young adulthood. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45*(7), 1288-1299.

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

Meredith, W., & Horn, J. L. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change.* (pp. 203-240): American Psychological Association.

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

Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, *26*(3), 479.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*(1), 63-75.

Muthén & Muthén. (2004). Mplus (Version 3.11). Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., & Khoo, S. T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and individual differences*, *10*(2), 73-101.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide*. 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.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*(3), 343-355.

Quirk, K. J., Keith, T. Z., & Quirk, J. T. (2001). Employment during high school and student achievement: Longitudinal analysis of national data. *The Journal of Educational Research, 95*(1), 4-10.

R Development Core Team. (2004). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517-529.

Raudenbush, S., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2 ed.). Thousand Oaks, CA: Sage Publications.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329.

Raykov, T. (1998). Coefficient Alpha and Composite Reliability with Interrelated Nonhomogeneous Items. *Applied Psychological Measurement*, *22*(4), 375-385.

Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*(1).

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Reuterberg, S.-E., & Gustafsson, J.-E. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement, 52*, 795-811.

Roth, D. L., Haley, W. E., Owen, J. E., Clay, O. J., & Goode, K. T. (2001). Latent growth models of the longitudinal effects of dementia caregiving: A comparison of african american and white family caregivers. *Psychology and Aging, 16*(3), 427.

Rovine, M. J., & Molenaar, P. C. M. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online, 3*(2).

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*(1), 131-151.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*(1), 83-90.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179-200). Washington, DC: American Psychological Association.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York, NY: Oxford University Press.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research, 95*(6), 323-332.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72-101.

Stoel, R. D., van den Wittenboer, G., & Hox, J. (2004). Including time-invariant covariates in the latent growth curve model. *Structural Equation Modeling, 11*(2), 155-167.

Tisak, J., & Meredith, W. (1990). Descriptive and associative developmental models. In A. v. Eye (Ed.), *Statistical methods in longitudinal research* (Vol. 2, pp. 387-406). San Diego, CA: Academic Press.

U.S. Department of Education. (2003). *No Child Left Behind, accountability and adequate yearly progress (AYP).* Paper presented at the National Title I Directors' Conference, Anaheim, CA.

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (1996). *Mathematical statistics with applications*. Belmont, CA: Duxbury Press.

Wallace, K. A., & Wheeler, A. J. (2002). Reliability generalization of the life satisfaction index. *Educational and Psychological Measurement, 62*(4), 674.

Willett, J. B., & Keiley, M. K. (2000). Using covariance structure analysis to model change over time. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 665-694). San Diego, CA: Academic Press.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*(2).

Willett, J. B., & Sayer, A. G. (1996). Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

VITA

Walter Lana Leite was born in Belo Horizonte, Brazil on February 26, 1978, the son of Nelson José Pereira Leite and Maria do Rosário de Lana Leite. After completing his work at Colégio Universitário, Viçosa, Brazil, in 1995, he entered the Universidade Federal de Minas Gerais. He obtained a bachelor degree in Psychology from the Universidade Federal de Minas Gerais in 2000. In August, 2001, he joined the graduate program in quantitative methods of the Department of Educational Psychology of the University of Texas at Austin. In August, 2003, he received a Masters of Arts in program evaluation from the University of Texas at Austin, and continued working towards his Ph.D. His publications include:

Leite, W. L. & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement*, 65(1), 140-154.

Stapleton, L. M. & Leite, W. L. (in press). A Review of Syllabi for a Sample of Structural Equation Modeling Courses. *Structural Equation Modeling*.

Beretvas, S. N. Meyers, J. L. & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570 -589.

Permanent Address: Rua Pastor Muryllo Cassete 240, Bairro São Bernardo, Belo Horizonte, Minas Gerais, Brazil, cep: 31750-530

This dissertation was typed by the author.