14
Latent Growth Curve Models

Kristopher J. Preacher

Structural equation modeling (SEM) is one of the most flexible and commonly used tools in the statistical toolbox of the social scientist. *Latent growth curve modeling* (LGM), the subject of this chapter, is one application of SEM to the analysis of change. In LGM, repeated measures of a variable (hereafter, *Y*) are treated as indicators of latent variables, called *basis curves*, that represent aspects of change—typically intercept and linear slope factors. Values of the time metric (e.g., age, day, or wave of measurement) are built into the factor loading matrix to reflect the form of the hypothesized *trajectory*, or trend over time. There are many extensions of this idea, but these are the basic elements common to all applications of LGM. LGM contains elements of both variable-centered and person-centered approaches (Curran & Willoughby, 2003), in that a sample-level summary of change is provided, yet individual differences in initial status and change are also considered.

The advantages of LGM over rival techniques for modeling change are numerous. A primary advantage is that LGM affords the researcher the ability to model aspects of change as *random effects*; that is, the means, variances, and covariances of individual differences in intercepts and slopes can be estimated. Because LGM is a special case of SEM, all of the benefits of SEM apply to LGM as well. These include the ability to assess the fit of the model to data, the ability to assess change in latent variables, and the ability to examine the antecedents and sequelae of change. Missing data (assuming they are missing at random) pose no problem for LGM. Perhaps the greatest advantage of LGM is its flexibility. Cases need not be measured at the same occasions, or even at equally spaced intervals. Complex nonlinear trajectories can be modeled. LGM can be adapted in creative ways to address new problems.

There are currently four book references devoted exclusively or primarily to LGM (Bollen & Curran, 2006; Duncan, Duncan, & Strycker, 2006; Preacher, Wichman, MacCallum, & Briggs, 2008). In addition, there exist many accessible articles and chapters on the subject (e.g., Byrne & Crombie, 2003; Chan, 1998; Curran, 2000; Curran & Hussong, 2003; Hancock & Lawrence, 2006; Singer & Willett, 2003; Willett & Sayer, 1996). Any SEM software capable of accommodating mean structures and multiple groups (e.g., AMOS, EQS, LISREL, Mplus, Mx) may be used to specify these models.

Because LGM is a special application of SEM, the reader may notice some degree of overlap between the desiderata enumerated here and those described in Chapter 28 of this volume. Table 14.1 of the present chapter addresses some of these in the context of LGM, and includes *additional* desiderata specific to the case of LGM. To get the most out of this chapter, it should be read in conjunction with Chapter 28.

Table 14.1 Desiderata for Latent Growth Curve Models

Desideratum	*Manuscript Section(s)**
1. Substantive theories motivating the model under scrutiny are described; a set of a priori specified competing models is generally preferred.	I
2. The metric of time (or, more generally, the substrate of change) should be described.	I
3. The functional form of the hypothesized trajectory of change is described.	I
4. Path diagrams are presented to facilitate the understanding of the conceptual model of change and the specification of the statistical model.	I
5. The scope of the study is outlined; if the study delineates a theory of change over time, enough time must be permitted to elapse for the phenomenon of interest to unfold.	I
6. Repeatedly measured variables are defined and their appropriateness for inclusion in the study is justified.	M
7. How any theoretically relevant control variables are integrated into the model is explained.	M
8. The sampling method(s) and sample size(s) are explicated and justified.	M
9. The treatment of missing data and outliers is addressed.	M, R
10. The name and version of the utilized software package are reported; the parameter estimation method is justified and its underlying assumptions are addressed.	M, R
11. Problems with model convergence, offending estimates, and/or model identification are reported and discussed.	R
12. Summary statistics for measured variables are presented; if raw data were analyzed, information on how to gain access to data is provided.	R
13. Recommended data-model fit indices from multiple classes are presented and evaluated using literature-based criteria.	R
14. If incremental fit indices are used, an appropriate null model is specified and fit rather than relying on incorrect estimates provided by software.	R
15. For competing models, comparisons are made using statistical tests (for nested models) or information criteria (for non-nested models).	R
16. For any post hoc model respecification, theoretical and statistical justifications are provided and the model is fit to a new sample.	R
17. Parameter estimates, together with information regarding their statistical significance, are provided.	R
18. Appropriate language regarding model tenability and structural relations is used.	D

* *Note*: I = Introduction, M = Methods, R = Results, D = Discussion

1. Substantive Theories and Latent Growth Curve Models

LGM is intended as a way to test the a priori predictions of a theory of change against observed data. Therefore, it is critical that the researcher have a well-articulated theory of change before attempting to use LGM. The Introduction section of an article using LGM should build a case for testing specific hypotheses of change. Typically this involves stating a theoretical reason for specifying individual trajectories that are characterized by aspects of change (intercept, slope, and so on) that, in turn, are expected to vary across sampling units. The point of most applications of LGM is to obtain estimates of the means, variances, and covariances of these trajectories, and these parameters should have consequences for the theory under scrutiny. The lack of strong theoretical predictions can lead to the misuse of LGM to generate theory from data in an exploratory, inductive fashion. As with SEM in general, testing alternative models of change provides a more comprehensive survey of competing ideas—and is more scientifically sound—than testing a single model of change.

Given that the researcher has in mind a strong theory of change that makes LGM an appropriate analytic strategy, then attention must be given to the question "change in what?" Most applications of LGM involve modeling change in the same variable over time, but this is a questionable undertaking if the nature or meaning of the variable itself changes over time. Change in the fundamental meaning of the variable could be mistaken for change in mean level. For example, common age-appropriate aggressive behaviors in 6-year-old children may decrease in frequency over time not because levels of the aggression construct decline, but rather because other aggressive behaviors take their place. One way to address this problem is by invoking theory or past findings to support the stability in interpretation of the repeatedly measured variable. Another way is to replace directly observed repeated measures with repeated latent variables, each of which has multiple indicators at each measurement occasion. This approach permits tests of *longitudinal factorial invariance*, a way to assess stability in the meaning of a construct over time. Applications that use repeated latent variables should formally address longitudinal invariance.

2. The Metric of Time

The overwhelming majority of applications of LGM involve some metric of time as the substrate of change. Time can be measured in units ranging from milliseconds to decades. Quasi-time metrics such as wave of study, developmental stage, or school grade may be used. In fact, the data analyzed with LGM need not be longitudinal at all. For example, there is theoretically no hindrance to replacing the time metric with, for example, stimulus intensity, distance, or dosage, assuming these repeated measures are assessed or administered in a within-subject fashion and ordered in some logical way. In the case of stimulus intensity, the "origin" measure (typically time 0 in a longitudinal study) could represent the absence of a stimulus; it could represent the baseline dosage of a drug in a repeated-measures medical trial. For the remainder of this chapter I refer to the metric as "time," but this is not meant to exclude other substrates of change.

Two factors should be considered in any application of LGM: *origin* and *scale*. The origin of the time metric refers to the zero-point. The location of the zero-point (e.g., age 37, initial wave of measurement, or time of intervention) has implications for the interpretation of model parameters related to the intercept. For example, the first occasion of measurement is often chosen as the origin to permit interpretation of the intercept as "initial status," although other choices are feasible and more appropriate in different circumstances (e.g., "time of death" as the last measurement occasion). The scale of a metric refers to the unit of time (e.g., year, minutes since treatment, or developmental stage). Scale is important mainly for interpreting parameters related to the slope, because (linear) slopes are interpreted as the model-implied change in the outcome per unit increase in time. The choice of origin and scale either should be justified by the researcher or should be obvious from the context, and should not be chosen arbitrarily.

3. Functional Form

Most applications of LGM involve testing linear trends. That is, the researcher hypothesizes that scores on the repeated measures proceed upward or downward in a linear fashion. Individuals may vary around this mean linear trend if intercept and slope variances are also estimated, meaning that the basic growth curve model provides for variability in level and rate of change. In LGM, values representing the trend are entered into columns of a matrix of factor loadings ($\mathbf{\Lambda}$) in the following way. The first column of $\mathbf{\Lambda}$ always consists of a column of 1s to act as multipliers for the intercept factor. The remaining columns represent functions of the values of the time metric. For example, for a simple linear trend with four equally spaced repeated measurements, $\boldsymbol{\Lambda}$ could be represented in any of the following equivalent ways:

$$\boldsymbol{\Lambda}_{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \boldsymbol{\Lambda}_{B} = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \quad \boldsymbol{\Lambda}_{C} = \begin{bmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \quad \boldsymbol{\Lambda}_{D} = \begin{bmatrix} 1 & 10 \\ 1 & 20 \\ 1 & 30 \\ 1 & 40 \end{bmatrix},$$

where the first column is always the constant 1 and the second column (containing actual values of the time metric) represents linear growth. The factor loadings in the $\boldsymbol{\Lambda}$ matrix often are represented in diagram form as path values on arrows connecting intercept and slope factors to the repeated measures of the outcome variable. In Figure 14.1, the loadings in $\boldsymbol{\Lambda}_A$, $\boldsymbol{\Lambda}_B$, $\boldsymbol{\Lambda}_C$, and $\boldsymbol{\Lambda}_D$ are depicted in simplified path diagram form (path diagrams are treated at greater length under Desideratum 4). Notice that the location of the zero-point, and thus the occasion at which the intercept is interpreted, changes from specification to specification, as does the metric of time. The choice of both the origin and scale of the time metric should be consistent with theory and with the research context. The origin should be chosen carefully to correspond with a theoretically important occasion—for example, the time of initial assessment, time of intervention, or time of death—so that time can be thought of as *time since* (or until) that event. In the first loading matrix ($\boldsymbol{\Lambda}_A$), the origin is placed at the first occasion of measurement. In the second ($\boldsymbol{\Lambda}_B$), the origin occurs at the second occasion of measurement, and so on. The scale is always chosen to correspond to a theoretically important metric. For example, change over time in $\boldsymbol{\Lambda}_B$ might be measured in two-month intervals, with an intervention imposed at the second occasion of measurement, whereas in $\boldsymbol{\Lambda}_D$ the unit of time is age in years, and the origin (i.e., birth) falls 10 years before the first assessment. It is usually not necessary to explicitly provide the $\boldsymbol{\Lambda}$ matrix, especially if an appropriately labeled path diagram is provided (see Desideratum 4), but it often can be helpful as an aid to understanding.

The basic linear LGM can be extended in numerous creative ways. For example, the first $\boldsymbol{\Lambda}$ matrix below ($\boldsymbol{\Lambda}_E$) specifies quadratic growth for four equally spaced repeated measurements—the first column provides multipliers for an intercept factor, the second for a linear factor, and the third for a quadratic factor. The $\boldsymbol{\Lambda}_F$ loading matrix provides for linear growth over four unequally spaced

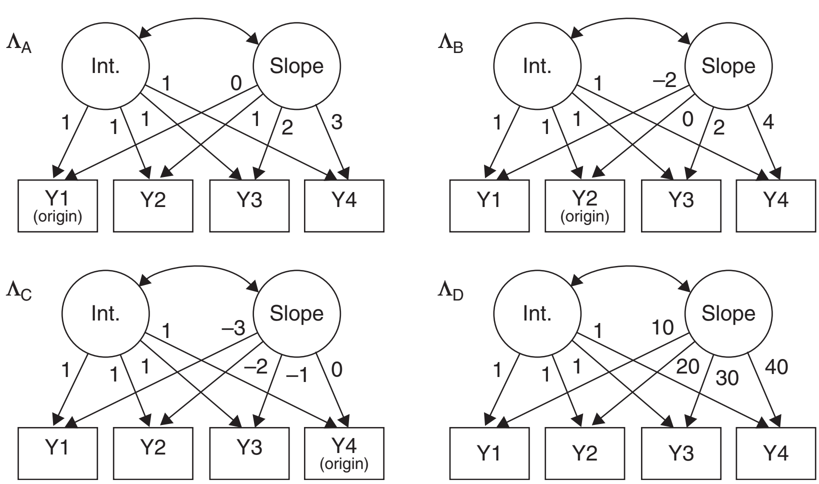

Figure 14.1 How the Loadings in $\boldsymbol{\Lambda}_A$, $\boldsymbol{\Lambda}_B$, $\boldsymbol{\Lambda}_C$, and $\boldsymbol{\Lambda}_D$ Might Be Represented in Path Diagrams.

measurement occasions (the second occasion being 4 time units since the initial occasion, the third 5 units, and the fourth 7 units). The $\mathbf{\Lambda}_G$ loading matrix represents an unspecified trajectory, in which the researcher has no specific trajectory in mind, but is willing to let the data determine the shape of change over time.

$$\mathbf{\Lambda}_E = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \quad \mathbf{\Lambda}_F = \begin{bmatrix} 1 & 0 \\ 1 & 4 \\ 1 & 5 \\ 1 & 7 \end{bmatrix} \quad \mathbf{\Lambda}_G = \begin{bmatrix} 1 & 0 \\ 1 & \lambda_{2,2} \\ 1 & \lambda_{3,2} \\ 1 & 1 \end{bmatrix}$$

There are two major points that should be addressed concerning functional form. The first is that neither the functional form of the hypothesized trajectory nor the measurement schedule need to conform to a rigid and limited set of options. Flexibility is a hallmark of LGM. The second point is that, regardless of what trend is hypothesized and fit, it must be explicitly justified on the basis of theory. It is rarely a good idea to use LGM in a theoretical vacuum, or to use it to approximate a trend of unknown shape for descriptive purposes. The option to approximate a functional form rather than test an a priori hypothesized one does exist, as the $\mathbf{\Lambda}_G$ matrix above demonstrates, but this practice is exploratory, not confirmatory, so conclusions should be worded to reflect the partly atheoretical nature of such trends.

4. Path Diagrams

The use of *path diagrams* is explained in Chapter 28 of this volume. Everything said about path diagrams in Chapter 28 applies here as well, because LGM is a special case of SEM. Path diagrams are not

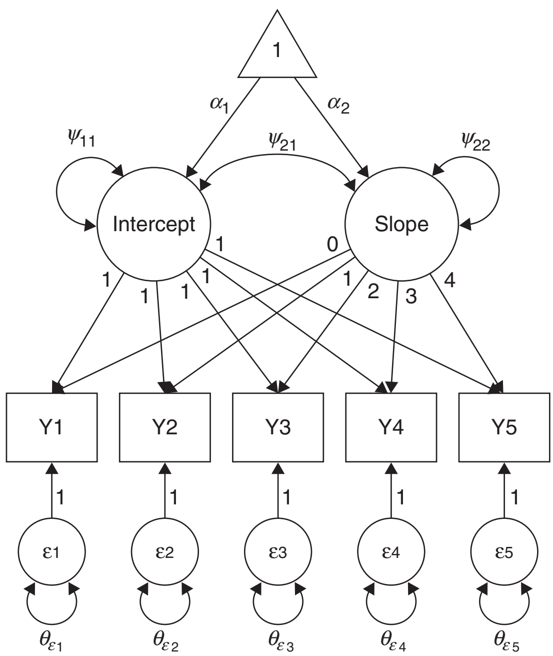

Figure 14.2 Path Diagram of a Linear Latent Growth Curve Model with Random Intercepts, Random Slopes, and Unconstrained Residual Variances.

required in applications of LGM, but they almost always greatly facilitate interpretation, especially for readers unacquainted with the method.

Very spartan diagrams were used to illustrate loadings under Desideratum 3. An example of a full latent growth curve path diagram is given in Figure 14.2. As in other SEM path diagrams, circles represent latent variables, squares are measured variables (here, repeated measures), single-headed arrows are path coefficients (regression-type weights), and double-headed arrows are variances or covariances. Aspects of change (intercepts, slopes, and so on) are considered latent variables because they cannot be directly observed. They usually are permitted to vary across people and to covary with one another (e.g., initial status may covary with rate of change), so parameters representing those variances and covariances are often included in the diagram (therein labeled as ψ). Together, the Intercept and Slope factors comprise the *latent trajectory*. In addition to the information provided in Chapter 28 in this volume, there are several noteworthy features specific to diagrams used in LGM. The triangle represents a constant 1.0, and is otherwise treated as a variable. Therefore, the path coefficients labeled as α_1 and α_2 represent the means of the Intercept and Slope latent variables, respectively. Occasion-specific residual variances, labeled as $\theta_{\varepsilon(1-5)}$ in Figure 14.2, are included to represent the portion of the variance in the outcome not explained by the latent trajectory.

The other noteworthy feature of Figure 14.2 is the set of loadings connecting the latent trajectory factors with the outcomes. Unlike most applications of SEM or confirmatory factor analysis (see

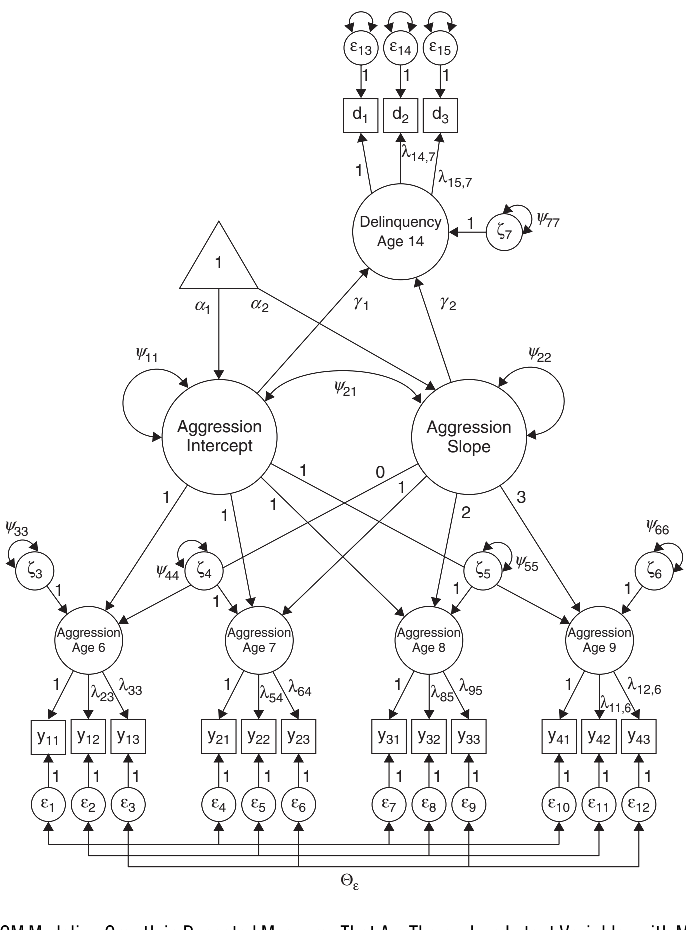

Figure 14.3 A Linear LGM Modeling Growth in Repeated Measures That Are Themselves Latent Variables with Multiple Indicators.

Chapter 8), all of these factors are typically fixed to point values that reflect the hypothesized trajectory. In Figure 14.2, that trajectory is linear, but that can be changed by altering the elements of the $\mathbf{\Lambda}$ matrix (see Desideratum 3). If any elements of $\mathbf{\Lambda}$ are freely estimated, the researcher is approximating an unknown trend rather than testing a hypothesis about a specific trend.

Growth curve models do not have to conform to the linear model depicted in Figure 14.2. For example, more latent trajectory factors may be added (e.g., quadratic, cubic). The equality constraint on residual variances can be freed, within certain limits (see Desiderata 5 and 11). The observed repeated measures can be replaced with latent variables. Because LGM is a special application of SEM, Intercept and Slope factors may serve as independent or dependent variables in larger path diagrams. Multiple latent growth curves may be included in the same model. Figure 14.3 depicts an elaborate example of hypothesized linear growth in a latent aggression construct from ages 6 through 9, where the effects of Intercept and Slope on Delinquency at age 14 are of interest. Residuals of similar indicators (i.e., those of y_{11}, y_{21}, y_{31}, and y_{41}) are permitted to covary in this example.

5. Scope of the Study

Scope in this context refers to the number and range of repeated measurements. These models find their greatest use when there are only a few repeated measurements per case (but usually at least 4 or 5), and when the sample size is large (see Desideratum 8), although LGM is by no means limited to such situations. The three most important questions to consider about the scope of a study are:

1. *Were a sufficient number of occasions chosen to identify the model?* A model is *identified* if the type and number of constraints is sufficient to guarantee a unique solution for every model parameter. For time-balanced data (in which data are collected over a limited number of discrete occasions), there must be at least $k + 1$ repeated measures for the model to be identified, where k is the number of basis curves (i.e., growth trajectory factors). This rule applies regardless of whether or not the residual variances are constrained to equality over time. For example, because a cubic trend requires four basis curves (intercept and linear, quadratic, and cubic slope components), there must be at least five repeated measures. If there are exactly five repeated measures and residual variances are constrained to equality, then $df = 5$. If residual variances are freely estimated, then $df = 1$. This rule is simply an elaboration of the assertion that two points define a line. At least two repeated measures are required in order for linear growth to be defined, but in order to test hypotheses about linearity, at least one more occasion is required. Identification rules become more complicated as the model departs from a basic growth curve. Fewer repeated measures may be required if some parameters are further constrained.

2. *Were a sufficient number of occasions chosen to adequately estimate the mean trend?* Even if the model is identified, $k + 1$ repeated measures are only barely enough to estimate parameters for a model with k basis curves. More than the bare minimum are generally preferred. Three occasions are required to test a hypothesis of linearity, but six occasions provide a far superior test of linearity. A large number of repeated measures becomes particularly important as the complexity of the hypothesized trend increases. It is also important to concentrate measurements around the part of the trend likely to show the greatest changes. For example, if the hypothesized trend is a negative quadratic curve (in which levels of Y are expect to start low, reach a maximum, and then decrease again), and the researcher has control over when measurements are scheduled, it is sensible to time the measurements such that they are more highly concentrated in the region of the maximum than in the extremes (although these are also necessary). Doing so generally reduces the likelihood of encountering estimation problems.

3. *Does the study span an interval sufficiently long to capture the process?* It is important to ensure that enough time elapses over the course of the study to adequately capture the process of interest. For

example, if growth in the height of elementary school children is measured on seven occasions spanning five weeks, clearly not enough time has elapsed to observe meaningful growth, and error of measurement likely will eclipse actual growth. If the trend under study is actually S-shaped (e.g., a learning curve) but data are collected during only a brief window of this process, the trend may appear to be linear. In such cases, a linear LGM might provide spuriously good fit, whereas a more appropriate nonlinear LGM may appear to overfit the data. By the same token, extending measurement beyond what is necessary to capture a trend can waste resources. If a simple linear trend is assessed across 18 occasions, the researcher may be better served to collect data from more cases at fewer occasions.

The key points regarding the scope of the study are that the number, spacing, and range of repeated measurements all need to be considered and justified on the basis of theory and minimum identification requirements. The number of repeated measurements should exceed the minimum, but should not be so numerous as to be wasteful. Measurements should be concentrated around regions of greatest expected change. Finally, the interval spanned by the first and last measurement must be sufficient to permit the process under study to unfold.

6. Repeatedly Measured Variables

As in any application of SEM, all measured variables in the model should be defined clearly, or else references should be provided. Some basic requirements of the repeatedly measured variables are that they be reliable and valid, and they should represent some attribute or characteristic that is able to change in *level* over time, but not *meaning*. The repeatedly measured Y should not represent stable attributes for which there is no theory of systematic change. In addition, it is necessary to provide a theoretical rationale for expecting not only growth, but growth of a particular form in a particular direction. If intercept and slopes factors are expected to vary across individuals, this expectation should have its roots in theory. In short, it should be realistic and theoretically appropriate to expect change in the chosen Y variable, and Y should be demonstrably reliable and valid.

As in other applications of SEM, special estimation procedures are required for variables that are ordinal, binary, or censored. Application of standard LGM to discontinuous or nonnormally distributed variables violates key assumptions necessary for legitimate statistical inference. Special procedures must be invoked.

7. Control Variables

Often researchers may want to examine the effects of some variables on others after controlling for covariates. Chapter 28 in this volume outlines some rules that should be followed in including covariates in SEM, and these rules largely apply in LGM as well. In LGM, two kinds of covariates are distinguished based on their location in the model. *Time-varying covariates* (TVCs), as their name implies, are predictor variables modeled at the level of the repeated measurements. They may be included in the model either as distinct variables predicting Y at each occasion of measurement, or as additional loading columns in the $\mathbf{\Lambda}$ matrix in the same way that the variable *time* is included in the $\mathbf{\Lambda}$ matrix. The first method is traditional practice in applications of LGM. The second is identical to how TVCs are included in hierarchical linear models (see Chapter 10, this volume). The LGM framework is flexible enough to permit either specification. Using the first approach, it is probably wise to permit TVCs to covary with other exogenous variables in the model (e.g., the latent trajectory factors).

Time-invariant covariates (TICs), on the other hand, are included at the subject level. TICs are exogenous predictor variables used to predict individual differences in aspects of change. For example, in Figure 14.3 we might introduce gender as a predictor of the Intercept and Slope factors. For both kinds of covariate (TVC or TIC), interest may be in controlling for the effects of covariates, that is, removing them from consideration so that effects of more substantive interest may interpreted more "purely." Or, interest may be in interpreting the effects of the covariates directly.

8. Sampling Method and Sample Size

The requirements regarding sampling method and sample size stated in Chapter 28 still hold when the model in question is a latent growth curve. If stratified or cluster sampling is used, it is crucial that this be considered in the modeling stage by employing sampling weights or multilevel SEM, or else the researcher should justify why these advanced procedures are not employed.

It is also essential that studies reporting latent growth curve analyses explicitly address the issue of sample size. As in non-LGM applications of SEM, there are several things to consider when choosing a sample size. First, *N* needs to be large enough to support the estimation of potentially many free model parameters. Maximum likelihood (ML) is a large-sample technique, and alternative estimation algorithms may require sample sizes much larger than ML. Second, the sample must be large enough to achieve adequate power for rejecting poor models by some criterion of fit. The criterion most commonly used for this purpose is RMSEA. Applications of LGM tend to have high power, but this does not release the researcher from the obligation to demonstrate that power is adequate in a given application. Third, the sample must be large enough for parameters of interest to have small standard errors (and thus narrow confidence intervals and high power). This is a very important consideration in LGM, where interpretation of parameters is of central interest. Finally, longitudinal studies are typically characterized by missing data due to attrition, death, late entry into the study, and other causes. The total sample size must be large enough to accommodate the amount of missing data. Although missing data techniques such as imputation may be used to fill the gaps in a data set to permit easy analysis, they cannot create information lost to attrition.

Ideally, the researcher should not only meet the minimum sample size suggested by these considerations, but exceed it by a considerable margin. The sample size may be just large enough to exceed the minimum required to achieve adequate power for tests of individual parameters, yet still fall short of the *N* required to yield usefully narrow confidence intervals. If the sample size is beyond the researcher's control, such as when samples of convenience or archived data are used for analysis, the researcher should still demonstrate that *N* is large enough to support estimation of a growth curve model and valid interpretation and testing of parameters. The important point here is the sample size should be justified on reasoned grounds, not simply reported.

9. Missing Data and Outliers

Very few studies have complete data on all variables for all cases. In long-term longitudinal studies, where the same individuals are followed over time, there are particularly many reasons some observations may be missing for some cases, and missing data in turn may threaten the generalizability of results. These reasons may include attrition due to illness, incarceration, death, data management errors, late entry into the study, lack of subject motivation to follow through, and cancellation of research funding. Late entry and attrition are particularly dangerous in LGM studies, as data missing due to late entry or attrition are typically *not* missing at random. Longitudinal studies with nontrivial amounts of missing data at the beginning or end of the studied trajectory are subject to severe bias in intercept and slope mean and variance estimates.

Four broad strategies for dealing with missing data may be identified: *prevention*, *deletion*, *full-information*, and *imputation*. The best strategy to address missing data is to preemptively prevent data loss by design. Researchers should make every reasonable effort to minimize the proportion of missing data. Data deletion strategies (i.e., pairwise and listwise deletion) use only those cases with complete data for all or some of the variables, and are usually to be avoided. Full-information strategies involve estimating model parameters using all available information, even if some of that information comes from cases with incomplete data. Usable information can be gleaned even from cases with a single valid data point. Imputation strategies use information from existing data about the relationships among variables, and then fill in, or *impute*, reasonable values for the missing data. Complete-case methods are then applied to the imputed data set. Mean imputation, in which missing values of a variable are replaced with the sample mean, should be avoided because it often results in a distribution with unrealistically many cases at the mean value. If imputation is used, multiple imputation (in which several data sets are imputed and the results are averaged across imputations) usually gives the best results. There are other, less often used strategies but these account for the majority of them. Full-information and multiple imputation methods are those most often recommended by methodologists to deal with missing data. Pairwise and listwise deletion and mean imputation should not be used without extraordinarily compelling reasons.

In LGM, data may be missing by design. For example, in research that combines multiple overlapping cohorts, one cohort may be measured at occasions 1, 2, 3, and 4, while "missing" the fifth occasion of measurement, whereas another may be measured at occasions 2, 3, 4, and 5, while "missing" the first occasion of measurement. Combined, all five occasions are represented, but the first and fifth occasions are not as well represented as the other three. This kind of data collection strategy can save time, but it obliges the researcher to assume that it is legitimate to combine cohorts to form a single trajectory. This assumption can and should be tested rather than assumed. If multiple cohorts are included, data missing by design from one cohort should not be imputed because it can lead to the creation of extrapolated data that are unrealistically congruent with those from other cohorts.

In summary, some general guidelines for reviewing studies with missing data can be developed. The amount and kind of missing data should be explicitly addressed, as should the likely reasons missing data were missing, the steps taken to address missingness in the analysis, and the likely impact of missing data on statistical analyses, the study's conclusions, and generalizability. Full-information and multiple imputation strategies are usually the best choices for dealing with missing data. Regardless of the strategy chosen to address missing data, the researcher should justify that choice. In reporting missing data, it is helpful to report the percentage of missing data for each variable at each occasion. Reporting only the percentage of complete cases does not provide enough information.

10. Software and Estimation Method

For key software considerations the reader is referred to Chapter 28 in this volume. Not all SEM software packages can fit growth curve models. Because the key parameters in LGM include the means of latent variables, the software must be capable of modeling means. Currently, the major SEM software packages capable of employing LGM are AMOS, EQS, LISREL, Mplus, Mx, and OpenMx. Because SEM software is regularly updated and improved, it is important to list the name and version of the software used to fit models and obtain parameter estimates.

11. Problems with Convergence, Estimates, and Identification

All of the advice in the corresponding section of Chapter 28 applies to LGM. Errors of identification, convergence, and estimation occur routinely in specifying and fitting models. Accurate documentation of these problems, and the steps taken to remedy them, is essential in reporting results.

It takes a fair amount of skill to properly specify a standard structural equation model, but LGM often requires even greater facility with software and knowledge of the mathematics behind the model. It is easy to make mistakes in specifying a model, and often these mistakes go unnoticed because the software is incapable of distinguishing sensible models from nonsensical ones, or because software will cheerfully provide reasonable-looking results despite serious problems. Here three examples of problems that sometimes occur in the application of LGM, and which may go unnoticed by inexperienced researchers, are described.

First, if a model is underidentified, some SEM software applications (e.g., LISREL and Mplus) may automatically and unobtrusively add constraints to some parameters to render the model identified. Occasionally the researcher is unaware that this automatic identification occurs, and parameters that should not be interpreted are interpreted nevertheless.

Second, negative or boundary values of variance parameters may be reported. Latent growth curve models are fairly robust to estimation problems, but if the model is severely misspecified some very strange things may happen. For example, residual variances may sometimes be estimated as negative (or, depending on the software, constrained to a boundary value of zero). This kind of result is a serious and common estimation error known as a *Heywood case*, and is usually indicative of serious model misspecification or, sometimes, a sample that is too small. If a residual variance of zero is reported, it is likely a Heywood case rather than a true zero. This kind of error may go unnoticed and unreported because the parameters of central interest in LGM are those related to the intercept and slope factors, not typically the residuals. Thus, if residual variances are not reported, they should be.

Third, covariances among aspects of change (intercept, slopes) may correspond to correlations that lie outside of the logical bounds of –1.0 and +1.0. This problem may not be immediately recognizable if only variances and covariances are reported. For example, both covariance matrices below, labeled as $\boldsymbol{\Psi}_A$ and $\boldsymbol{\Psi}_B$, may appear to be legitimate covariance matrices at first glance, but only $\boldsymbol{\Psi}_A$ is acceptable or "proper." The covariance in $\boldsymbol{\Psi}_A$ equates to a reasonable correlation of $.28/(.19^{1/2} \times .64^{1/2}) = .80$, but the covariance in $\boldsymbol{\Psi}_B$ equates to an impossible correlation of $.38/(.19^{1/2} \times .64^{1/2}) = 1.09$.

$$\boldsymbol{\Psi}_A = \begin{bmatrix} .19 & .28 \\ .28 & .64 \end{bmatrix} \qquad \boldsymbol{\Psi}_B = \begin{bmatrix} .19 & .38 \\ .38 & .64 \end{bmatrix}.$$

If an improper matrix is reported, this indicates that an undetected estimation error has probably occurred. Such errors sometimes can be addressed by correcting coding errors, removing outliers from the data, or providing better starting values for the estimation procedure. Sometimes the problem cannot be solved, which usually indicates that specified growth model is not appropriate for the data.

12. Data Display and Accessibility

Fitting latent growth curve models requires access either to raw data or to a covariance matrix and mean vector. In line with recommendations endorsed in Chapter 28 of this volume, the data should be reported or made available to permit other researchers to verify or reexamine reported results. Whenever possible—given the constraints of the study, journal space, and proprietary issues—summary information in the form of a covariance or correlation matrix, mean vector, and standard deviations for complete data typically are sufficient to permit reanalysis. If some data are missing or if journal space does not permit reporting summary data, authors should provide instructions informing readers how they may obtain the data.

13. Data-Model Fit

A primary advantage of SEM is that it permits the assessment of fit between the model and data. The advice offered in Chapter 28 of this volume is reiterated here. The χ^2 statistic by itself has only limited usefulness as a fit index because of its sensitivity to sample size and trivial departures from perfect fit. It is usually wise to report multiple (at least three) fit indices drawing from the three broad classes (*absolute indices*, *parsimonious indices*, and *incremental indices*). SRMR, RMSEA, and TLI (NNFI) are perhaps the best representatives of these classes, but individual researchers may choose not to limit themselves to these, or to include all of them. If RMSEA is reported, its associated 90% confidence interval should also be reported and interpreted. Models may also be compared on the basis of relative fit.

14. Appropriate Null Model

Incremental fit indices like those discussed in Chapter 28 (NFI, NNFI/TLI, and CFI) constitute an important class of fit indices. They express the fit of a substantive model as falling somewhere between the fit of a highly restrictive "null" model and that of a saturated (perfectly fitting) model. The appropriate null model must satisfy two requirements: (a) it must be nested within the most restrictive substantive model to be tested and (b) it must constrain all covariances among observed variables to zero (Widaman & Thompson, 2003). The null model typically used in SEM is one in which only the means and variances of the observed variables are estimated, constraining the covariances to zero. In ordinary applications of SEM, this null model is perfectly appropriate. LGM, however, requires a different null model. Specifically, if the model to be tested is any growth curve model in which the residual variances are all freely estimated, the appropriate null model is an intercept-only model in which the only free parameters are the intercept mean and the free residual variances—a model with $p + 1$ free parameters that implicitly constrains the variables' means to equality but permits them to have different variances, and constrains their covariances to zero. On the other hand, if the model to be tested is any growth curve model in which the residual variances are constrained to equality, then the appropriate null model is an intercept-only model in which the only free parameters are the intercept mean and the constrained-equal residual variance (only two parameters). Thus, the incremental fit indices reported by default by most SEM programs are incorrect for latent growth curve models. They must be computed by hand by explicitly fitting the appropriate null model, obtaining the resulting χ^2 statistic, and manually computing the desired incremental fit index.

15. Model Comparisons

Investigating alternative models of change is a sound and recommended approach to theory evaluation. Compared to the evaluation of individual models in isolation, the comparison of rival models of change has a better chance of eliminating some theories from consideration. For recommendations regarding model comparison, the reader is referred to Chapter 28, Desideratum 14. Model comparison proceeds exactly the same way in LGM as in the more general SEM. Examples of the kinds of models one might wish to compare in LGM could include linear vs. quadratic models, or models in which residual variances are constrained to equality vs. freely estimated.

16. Model Respecification

Latent growth curve models are notoriously poor-fitting by traditional criteria. This poor fit arises not because LGM is unrealistically restrictive, but rather because most other applications of SEM have relatively many free parameters and show unrealistically *good* fit. The trajectories specified in LGM are highly constrained and are not likely to arise by chance in nature. Like any model, growth curve models are merely approximations to reality, and cannot be expected to fit perfectly. However, because tradition and publication pressures have made good fit a necessary component of publishing applications of SEM, the researcher may be tempted to counter instances of poor fit by freeing parameters identified by modification indices (see Chapter 28, Desideratum 15) and fitting the modified model to the same data, resulting in better fit. This temptation should be resisted. A good rule of thumb is that a model may be modified to any degree on the basis of modification indices, but (a) the modifications must be theoretically appropriate and (b) the modified model should be fit to new data to avoid the possibility of capitalizing on chance characteristics of the sample.

Using modification indices is especially discouraged in LGM because relaxing constraints on the model can severely compromise the interpretation of the model as a specific trajectory. For example, if the model in Figure 14.2 were fit to data and the software reports a large modification index for the loading connecting *Y4* to the Slope factor, freeing the loading may improve fit, but the resulting model can no longer be interpreted as a linear growth curve. Furthermore, the model no longer represents a test of the original, theoretically prescribed hypothesis. Even when not fit to new data, modified models nevertheless have use as descriptive tools. The main point of this desideratum is that modified models may be useful, but tests of such models should not be treated as confirmatory or as strict tests of hypotheses about growth over time.

17. Parameter Estimates and Significance

The end product of model-fitting is a collection of parameter estimates and fit indices. Assuming that model fit is reasonable and that no convergence, estimation, or identification problems persist, it is important to report the magnitude and significance of all model parameters—not only those that are of central interest, but *all* of them. In typical applications of LGM this number is not large. In the basic linear LGM with homoscedastic residuals, for example, there are only six parameters to report: the mean intercept and slope, intercept and slope variances, their covariances, and a common residual variance. More complicated models result in more parameter estimates to report.

A simple way to report parameter estimates is to place the point estimates in the appropriate locations in a path diagram (see Desideratum 4) along with some indication of significance or precision, such as confidence intervals, standard errors, *p*-values, or a system of asterisks indicating levels of significance. The method by which significance is determined is an important and often overlooked consideration. For some parameters—path coefficients and latent variable means—the traditional *z*-tests (dividing the point estimate by its standard error) usually are appropriate. However, for variance and covariance parameters, *z*-tests are not appropriate because such tests require the assumption that the tested parameter is normally distributed over repeated sampling, an untenable assumption for variances. A better test for variances and covariances is the *likelihood ratio test* (see Chapter 28, Desideratum 14), in which the fit of a model that fixes the parameter to zero is compared to the fit of a model in which the parameter is freely estimated. Depending on the parameter's role in the model, this test may not always be possible or appropriate. Alternatives are to obtain bootstrap confidence intervals for parameter estimates (an option available in several SEM software applications) or likelihood-based confidence intervals (available in Mx).

18. Interpretive Language

Several important points from Chapter 28 regarding interpretive language are reiterated here. First, latent growth curve models, and the theories of change they represent, are never literally "true," nor can they ever be confirmed empirically. Such statements are extremely misleading and should be discouraged at every opportunity. Models represent formal hypotheses, and hypotheses may fail to be rejected for many reasons, among them low statistical power. In fact, any structural equation model (LGM included) can be made to fit perfectly by freeing a sufficient number of parameters, but perfect fit does not imply a confirmed model. The outcome variable *Y* is not observed as a continuous function of time, so there is no foolproof way to confirm that *Y* values corresponding to unobserved values of time would similarly conform to the trend followed by the observed values, even if perfect fit is observed in the sample. Similarly, growth curve models with zero or negative degrees of freedom would fit *any* data equally well (i.e., perfectly), regardless of what process generated the data. In neither case can perfect fit be taken as support for the researcher's theory of growth. Even identified models with good fit can be described only as tenable in light of the data.

Second, causal language should be used sparingly if at all. A theory's predictions may be causal in nature, but a model's results can never completely support a conclusion that a process is causal, regardless of how well-designed the study may be. Alternative explanations can always be devised. However, confidence that a process is causal may be strengthened by experimental manipulation (e.g., treatment vs. control), temporal precedence (causes must always precede effects in time), and strong theory that prohibits or limits plausible alternative explanations (e.g., kindergarten may "cause" growth in verbal knowledge, but never the reverse). In applications of LGM, the data are naturally longitudinal, but this does not grant a license to use causal language with impunity. At most, findings may be supportive of a causal process, but can never definitively demonstrate causality.

References

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

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics, 2*, 177–203.

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

Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research.* Mahwah, NJ: Erlbaum.

Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*, 526–544.

Curran, P. J., & Willoughby, M. J. (2003). Reconciling theoretical and statistical models of developmental processes. *Development and Psychopathology, 15*, 581–612.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Erlbaum.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 171–196). Greenwich, CT: Information Age Publishing, Inc.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling.* Thousand Oaks, CA: Sage.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*, 16–37.

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: Issues and techniques* (pp. 125–157). Mahwah, NJ: Erlbaum.