I. Introduction

Throughout the social and behavioral sciences, researchers have increased their reliance on longitudinal designs to address questions about the potential change relations among various dimensions of behavior, including personality. One problem that researchers face as they embark on a longitudinal study is the lack of a roadmap to help inform important design decisions. Particularly with longitudinal investigations, understanding the interface between research question, study design, and analytic technique is crucial. In this chapter, we offer a broad overview of the analytic techniques that can be used with longitudinal data and discuss important developmental design considerations. Our goal is to provide a guide for researchers that will help clarify the strengths and weaknesses of the various techniques and the types of answer that the techniques can provide.

For the most part, few developmental questions can be addressed with a single analytic approach because each approach is geared to address only certain aspects of the data. In our overview, we will describe the basic features of the techniques and discuss their merits with regard to the nature of the question they can address and the quality of the answer that they provide. In general, longitudinal techniques can be categorized into those that address changes in group means (e.g., repeated-measures ANOVA), those that address changes in the individual-differences standings among a sample of individuals (i.e., modeling between-person relations with procedures such as regression, path analysis, or structural equation modeling; SEM) and those that address changes intra-individually (i.e., modeling within-person change relations with procedures such as growth-curve and multi-level modeling).

In this chapter, we will focus primarily on the techniques derived from the general class of latent variable methods (i.e., SEM) for the analysis of change because of the various limitations...
with the classical (least-squares) general linear model approach. Specifically, the advent of SEM brought with it a number of features that make it an ideal technique for modeling longitudinal data. First, the problems of unreliability are remedied by the use of multiple indicators of latent constructs. Second, the inherent need to represent indirect paths is easily accommodated. Third, the problem of correlated \( \text{‘errors’} \) (i.e., item uniquenesses) is no longer seen as a violated assumption, but rather as effects that can and should be estimated. Fourth, classical approaches rely on a set of distributional assumptions related to normality or multivariate normality. While the most popular estimation procedure in SEM, maximum likelihood, makes similar assumptions, SEM programs have numerous ways of dealing with mild to moderate violations of the classical distributional assumptions (e.g., robust standard errors, bootstrapping capabilities, and alternative estimation methods such as weighted least squares). Finally, measurement invariance of the constructs across time, which is an assumption of classical methods, can be tested and evaluated in the SEM framework.

Although SEM procedures possess numerous advantages, some disadvantages of longitudinal SEM include the need for large sample sizes and the assumption that the sample is homogeneous with regard to the underlying change processes or mechanisms. Mixture-distribution modeling, however, attempts to address this latter assumption empirically (see below), or theoretically derived sub-groups (who are then assumed to be homogeneous with regard to the change process) can be modeled separately using the multiple-group capability of SEM procedures.

II. Latent Variable Longitudinal Methods

a. Two-wave Difference- and Residual-Score Approaches

The simplest form of a longitudinal study is a two-occasion measurement design. Although most developmentalists would argue that two waves do not sufficiently constitute a longitudinal design, the two-occasion approach can, in fact, inform developmental questions (Hertzog, Dixon, Hultsch, & MacDonald, 2003). However, the two-occasion approach does carry with it significant disadvantages for answering questions about change. Bereiter (1963) outlined three basic problems with using a simple difference score to indicate change: the paradoxical relationship between the test-retest correlation and the reliability of the change score, the negative correlation between initial status and the change score, and the lack of consistency in interpretability of the change score at different points in the distribution. In addition, when two-occasion data is analyzed using classical approaches (e.g., difference scores, gain scores, residual-change scores, repeated-measures ANOVA, etc.), quite restrictive assumptions must be met (e.g., sphericity of variances) and the effects of the ever-present problem of unreliability are significantly heightened (Rogosa & Willett, 1983). Problems related to measurement error include type II errors associated with elevated variance estimates, under-estimates of stability, and regression-to-(and-from)-the-mean effects – these problems can become exacerbated with time-dependent longitudinal data (McArdle & Nesselroade, 2002).

As mentioned, methodologists have remedied many of these problems by taking a latent-variable SEM approach. McArdle (2001), for example, has outlined how one can calculate a latent-difference or latent-residual score that does not suffer from the issues associated with measurement error or highly restrictive assumptions. Figure 1 provides a sketch of two alternative latent-variable approaches to modeling change scores across two measurement occasions. The key to the latent variable approach is that constructs are assessed by multiple indicators at each time point. This process produces variance in each latent construct that is error free and, as a result, the change in variance, \( \Delta \), is also measured without error.

As seen in Figure 1, both approaches decompose the variance of the Time 2 construct into
two components (note that the Time 2 disturbance factor is fixed at 0.0 to identify the model’s parameters as well as force the two-way decomposition). For the difference-score model (Figure 1, Panel A) the two components are (a) variance associated with one’s absolute standing at Time 1 and (b) variance associated with the absolute difference from Time 1. For the residual-score model (Figure 1, Panel B), the two components are (a) one’s relative standing at Time 1 (i.e., the degree of correlation) and (b) the change in relative standing at Time 2. Note that the graphic representation of the change scores in Figure 1 can also be modeled as depicted in Figure 2. Here, the disturbance factor represents the change variance but would also carry the indirect effects of Time 1 if predicting to an outcome construct.

Three features distinguish these two different approaches to representing change. First, both models are equivalent in terms of their ability to reproduce the observed variance-covariance matrix (although degrees of freedom differ). Second, because the difference-score model assumes a perfect linear relationship, the Time 1 construct and the difference at Time 2 will be correlated to some degree depending on how much the relationship between the two constructs is not a perfect 1.0 correlation. Third, because the residual-score model estimates the degree of relationship between the two time points, the Time 1 construct and the residual at Time 2 are orthogonal. Fourth, the mean of the difference factor in Figure 1, Panel A is the absolute group-wise change. This mean-level change is similar to what would be estimated by repeated measures ANOVA with the exception that the mean-difference is corrected for measurement error and does not assume homogeneity of the variances over time.

Given the differences in the information captured and modeled, the two approaches address different types of question. The latent-difference score is most useful to model mean change over time and the individual differences around that mean change, while the latent-residual score model is most useful to examine issues related to stability of individual differences and the individual differences around that stability. As we will see in the next section, the latent-residual score model forms the basis for standard individual differences models while the latent-change model forms the basis for recently proposed alternative individual differences models (McArdle & Bell, 2000; McArdle, 2001; McArdle & Nesselroade, 2002).

b. Standard Individual-Differences SEM Models

Figure 3 displays a simple cross-lagged panel design for assessing change. In this figure, we have extended the graphical representation to include 3 waves of data with two constructs at each wave. The two constructs at Time 1 are said to be exogenous because no other variables or constructs are assessed prior to them (although, this model can easily be expanded to include covariate influences that could predict to Time 1). The four constructs at Time 2 and 3 are endogenous constructs because they are assessed ‘down stream’ and the variance-covariance information among these constructs would have the different effects of the prior time points accounted for in the lagged regression estimation process. In this case, the relationship between the two constructs at Time 1 is interpreted as an exogenous covariance (or correlation, depending on the scaling method), while the within-time associations among the constructs at Time 2 and Time 3 are interpreted as endogenous or residual covariances (or residual correlations when standardized).

The lines that link the corresponding constructs between each time point are termed auto-regressions and account for the individual-differences stability of the constructs across each time point. If the change process is at a constant rate, the between-time stability coefficients are sufficient to capture the rate of change over time (i.e., no direct effects from Time 1 to Time 3 are required). Such an idealized pattern of change results in what is referred to as a simplex
covariance structure. If direct paths from Time 1 to Time 3 are required by the data, the change process no longer conforms to a simplex structure. A non-simplex structure indicates that one or more factors has influenced the change process making the association between Time 1 and Time 3 either stronger or weaker than would be expected if the change process progressed at a constant linear rate. Factors that can influence the rate of change include contextual/environmental influences or nonlinear maturational processes that speed up or slow down the rate of change.

The paths that cross-link the constructs over time are referred to as cross-lagged effects. These effects, when significant, indicate that change in one variable is related to prior status in the other variable. Because of the time-ordered relationships among the variables such effects have a ‘causal’ flavor about them, but cannot be considered causal because of the ever-present unmeasured variable problem and the exogeniety assumption. The unmeasured variable problem applies to any modeling endeavor and simply means that some other unmeasured variable may be the causal mechanism driving the observed pattern of influences. The exogeniety assumption also applies to any modeling endeavor and simply highlights the idea that the Time 1 constructs, which are assumed to be exogenous, are likely not representing the true beginning of the time-ordered sequence of relationships among the constructs.

This being said, the primary research questions that can be asked and answered by the standard individual-differences approach to longitudinal data include (a) are the constructs measurement invariant over time (i.e., are the measures tapping in to the same thing at different points in time)?, (b) how stable are the constructs over the observed time span (i.e., to what degree do the individual differences standings get shuffled over the time intervals assessed)?, (c) what are the relative mean-level differences in the constructs over time?, (d) is the change process adequately captured by a simplex process (i.e., is the rate of change linearly constant and unaffected by other sources of influence), (e) is there any evidence of cross-lagged influences that are predictive of the cross-time changes?, (f) are the cross-time changes reciprocal or predominantly unidirectional?, and (g) are the cross-time effects consistent between each adjacent time point?

In answering these basic questions about the change process, the modeling steps typically begin with a longitudinal CFA wherein all constructs are allowed to correlate with one another (i.e., no directed paths are estimated). This model is used to address and answer the first three questions – that is, the questions regarding (a) measurement invariance (which should include invariance of both the intercepts and loadings and allow all corresponding residuals to correlate over time; see Little & Slegers, in press; Meredith, 1993), (b) the general level of individual-differences stabilities over time (i.e., the estimated latent correlations from the CFA model), and (c) the relative mean-level changes over time.

The next step of the modeling process would then specify the expected pattern of cross-time influences (as depicted in Figure 3, for example). This model is typically more restricted than the CFA (i.e., fewer parameters are estimated) but should achieve a similar level of fit to be deemed adequate in capturing the change relationships over time. This model would be used to evaluate and answer the latter four questions regarding the possible simplex structure of the change process, the existence of cross-lagged effects, their relative strengths, and their consistency over time.

As can be seen, the standard individual-differences approach is useful for answering a number of relevant developmental questions (Burkholder & Harlow, 2003). Such models are also important to assess the adequacy and invariance of each construct’s indicators over time and
across groups (Little, 1997; Little, Lindenberger, & Nesselroade, 1999) and the potential for selectivity between continuers and dropouts (Little, Lindenberger, & Meier, 2000); however, the change dynamics are not fully examined in a standard individual-differences model. A complementary procedure for modeling change over time is to analyze the data using growth curve methodology.

c. Growth Curve Modeling

In this section, univariate and multivariate change models are presented as structural equation models and, in particular, latent growth curve models. These models are described in a variety of sources, including Aber and McArdle (1991), Duncan, Duncan, Strycker, Li, and Alpert (1999), McArdle (1988), Meredith and Tisak (1990), Raykov and Marcoulides (2000), and Willett and Sayer (1994). These models could easily be conducted using any of the modern SEM programs, such as Mplus, EQS, LISREL, Amos, Mx, or SAS PROC CALIS, and most of the models presented in this section can also be implemented using multilevel modeling software such as SAS PROC NLMIXED (see e.g., Ferrer, Hamagami, & McArdle, 2004).

Latent growth curve modeling is a procedure that allows one to model individual differences in the changes over time by implementing a random coefficient or multilevel framework within the SEM framework (Bauer, 2003; Curran, 2004). In a multilevel model, of which the hierarchical linear model is a type, a population of lower level units (level 1) are “nested” within a higher level population of units (level 2). In the traditional HLM classroom example, the population of students (level 1) are nested within classrooms (level 2). Nesting is necessary when clustering (and statistical non-independence) of observations occurs due to a higher-level influence. In this case, assessments of students within a classroom are considered to be correlated to some degree because of their shared within-classroom experience (see Bryk & Raudenbush, 1987, 1992; Raudenbush & Bryk, 2002; Hox, 2000).

Just as student assessments within a classroom may be correlated, repeated observations (level 1) on an individual student may be nested within the individual student (level 2). More specifically, growth curve models attempt to represent the dynamic nature of change as a “mixed” combination of fixed and random coefficients or effects, hence the term mixed model that is sometimes used to describe this type of analysis. In the mixed model, the basic components of change for a sample of individuals are (a) the fixed average intercept, or starting point; (b) the fixed average slope, or degree of change over time; (c) the random individual variability around the average starting point; and (d) the random individual variability around the average shape of change (Cudeck & du Toit, 2001). In the examples to follow, we make the key assumption that the sample is homogeneous with regard to the change process that is being modeled. That is, we assume that the data follows a 2-level hierarchy (observations within individuals) and there is not a higher level of sampling present in the data (i.e. a “level 3”) whether observed or latent. This assumption is also true of SEM in general. It is possible to conduct a 3-level multilevel growth curve analysis within an SEM framework (i.e. observations nested within persons and persons nested within schools – see Duncan, Duncan, Okut, Strycker, & Li, 2002; Hox, 2000, 2002; Heck & Thomas, 2000), but that type of analysis is beyond the scope of this chapter.

Latent growth modeling and multilevel modeling, as mentioned, can be specified such that they are identical approaches. However, modeling growth models in the SEM framework has two notable advantages over traditional multilevel approaches. First, latent growth modeling in the SEM context can easily be expanded to include distinct growth curves for multiple variables to examine the dynamic interplay among them (McArdle & Nesselroade, 2002). For example,
growth curves for each of the big five personality factors can be estimated simultaneously to examine the potential change dynamics among them. Second, in conjunction with the multiple-group processing capabilities of SEM programs, complex (i.e., non-monotonic) non-linear effects can be specified (McArdle & Nesselroade, 2002). Third, latent growth curve modeling allows for the modeling of change in a latent construct, with all the benefits related to latent variables versus manifest variables described earlier.

Turning back to the basic specification of a latent growth curve model, if the assumption of sample homogeneity is warranted, the specific nature of the change can be modeled in a number of ways. We will begin with the SEM alternative to the standard longitudinal mixed model assuming a linear trend over time and extend the model to more complex, and more interesting, specifications.

In Figure 4, we provide a series of alternate (unconditional) growth curve models that can be fit to four waves of data. Note that growth curve models can also be fit to just three waves of data, but the nature of the shape of change that can be modeled is more limited. For all the models depicted in Figure 4, the first construct represents the intercept or starting level of the curve that is being modeled. The loadings of each variable on this construct are all fixed at 1.0 representing the regression of the repeated outcome measures on a constant in order to capture the mean intercept, or starting point of the curve. The estimated variance of this construct reflects the individual variability around the grand mean and the estimated latent mean of this construct is the estimated value of the mean intercept. These parameters are equivalent to the random individual variability around the average starting point and the fixed average intercept, or starting point in the mixed model described above. The means of the latent constructs are represented in these diagrams as the regression of the constructs on to the unit constant ‘factor’ (e.g., Kappa and Alpha matrices of LISREL and V999 of EQS) which is depicted as a triangle in these diagrams (for more details of modeling means, see Browne & Arminger, 1995; Little & Slegers, in press; Little, Slegers, & Ledford, in press). The remaining latent constructs, labeled with an $S_i$ represent the change that occurs relative to the intercept and differs by model.

Simple polynomial models. We begin by illustrating a simple linear relationship between the dependent variable and time that provides a direct test of the actual linear slope of the growth trend using simple polynomial contrasts. A first polynomial model depicted in Figure 4, Panel A, is one in which the expected growth pattern is expected to be linear. In this model, all four loadings from the “shape” factor are fixed to values that reflect the proportional relationship between time points (e.g., 0, 1, 2, 3 or 0, 33, 66, 1 if the observations are equally spaced; or say 0, 1, 3, 5 or 0, 20, 60, 1 if the observations are unequally spaced). Note that the factor loading for the first variable is zero and denoted in Figure 4A as a dashed line. The estimates are said to be centered on the first measurement occasion with this pattern of specified loadings.

Specifically, the loading of the first variable is fixed at zero so that the intercept would take on the value of the starting point of the curve at the first occasion of measurement. If this value were fixed at 1, for example, the intercept would then take on the value of the starting point at some point prior to the first occasion of measurement (McArdle & Nesselroade, 2002). If the loading for the last variable were fixed at zero and the loadings for the other three variables were negative (e.g., -3, -2, -1, 0), the intercept would take on the value of the last measurement occasion and change would be interpreted as change leading up to a particular event rather than change since an event. See Kreft, de Leeuw, and Aiken (1995) or Rovine and Molenaar (1998) for additional discussions of centering.

The estimated variance of the shape construct reflects the individual variability around the
average linear change in the sample, and the estimated latent mean of this construct is an estimate of the average slope. These parameters are equivalent to the random individual variability around the average shape of change and the fixed average slope, or degree of change over time, in the mixed model described above. The remaining model parameters are the residual error terms for the four variables, which are commonly constrained to be non-zero but equal, and the covariance between the intercept and slope factors. It is reasonable to assume that the initial status on a variable is related to, or correlated with, the subsequent change or growth on that variable. While this assumption is commonly included in formulating a latent growth curve model, it is not always required (Raykov & Marcoulides, 2000). Other specifications might include a specific directional relationship between the level and shape factors such that the initial status on a variable directly impacts the nature of the subsequent change that occurs. As a result, instead of estimating the simple covariance between the level factor and the shape factor, a directional path may be modeled.

The residual terms are constrained to be equal under an assumption of homoscedasticity and are equivalent to the level 1 residual in a multilevel or mixed model. The equality constraint on the residual terms can easily be relaxed and subsequently reflect a heteroscedastic situation. The residual variances are estimated to account for the deviation of scores from the individual trend curves. The constraint of homoscedasticity (or stationarity) keeps the error impact equal at all time points and preserves model parsimony (Raykov & Marcoulides, 2000). A residual variance represents unexplainable or unaccountable error or individual differences in measurement. By assuming a common residual variance, one implies that the unexplainable variance at each time point is constant and that the remaining variance is due to the initial status of the trait measured by a particular variable and subsequent change in the trait over time.

This method is more parsimonious than the heteroscedastic case because it allows for additional degrees of freedom (i.e., fewer estimated parameters). For example, consider each of the variables measured at all four time points. Instead of estimating four parameters (one residual variance for each time point), it is only necessary to estimate one common residual variance parameter thus saving three model degrees of freedom. Within the context of SEM models, the assumption of homoscedasticity can be evaluated by examining both global and local model fit and relaxing the constraints if so warranted.

If the linear slope model fits the data well then one could conclude that the growth pattern is readily captured by a linear trend. If this linear model does not fit, however, then the linear trend is not warranted and a quadratic or more complex function may be a better representation of the change over time. Figure 4, Panel B, depicts a growth curve model to estimate a quadratic function. Here, the fixed loadings for the linear construct are the same as they were for the linear slope model, and the fixed loadings for the quadratic construct are the squares of the corresponding linear loadings (e.g. 0^2, 1^2, 2^2, 3^2) just as one might use “x^2” in addition to “x” to represent a quadratic relationship between predictor and criterion in regression. The variance of the intercept, or level (“L”), construct still represents the individual variability at the intercept, the variance of the linear slope construct (“S_L”) still represents the variability in change over the range of time, and the variance of the quadratic construct (“S_q”) represents variability in the acceleration or deceleration of the growth over time. The latent means of each of these constructs represent the average status at the intercept, the average growth over time, and the average acceleration or deceleration. Residual error terms are still included in the model as are covariances between latent factors.

Because there are four measurement periods in this example, the polynomial growth
function may be extended through a cubic component (i.e., the number of possible components is equal to the number of measurement periods minus 1).

*Optimal level and shape models.* The next two models in Figure 4 (Panel C and D) depict two alternative models that represent variations of the latent growth curve model previously presented. In both “level and shape” models, the second construct captures the shape of the change over time, but the shape of change over time is not specifically modeled as a particular function (i.e., linear, quadratic, etc.) as it was in the simple polynomial approach. Instead, this flexible approach to growth curve modeling allows for the parsimonious capture of non-linear change over time without the need for additional latent factors beyond the intercept (level) and shape factors, regardless of the order of curvature.

In both Panel C and Panel D of Figure 4, following the same consideration for centering discussed previously, the loading of the first variable is fixed at zero so that the intercept would take on the value of the starting point of the curve at the first occasion of measurement. As mentioned, if this value were fixed at 1, for example, the intercept would then take on the value of the starting point at some point prior to the first occasion of measurement (McArdle & Nesselroade, 2002). In order to set the scale or change units of the shape of change, a second loading must be fixed to some non-zero value. In Panel C, the loading of the second variable on the shape construct is fixed at 1.0. The loadings for the variables at Time 3 and Time 4 are then freely estimated to capture the unrestricted shape of the change over time. The metric of the loadings at Time 3 and Time 4 are standardized as proportional change units relative to the magnitude of the change between the first and second measurement occasion.

In Panel D, the loading of the last variable on the shape construct is fixed at 1.0, and the middle two time points are estimated. Here, the values of the middle two time points would be proportional to the overall change between Time 1 and Time 4. Both of these models are identical in terms of their fit. The only difference is in the scaling of the estimates (e.g., the loadings, the mean of the shape construct, and the variability around the estimated slope). The model in Panel D provides roughly standardized values, so it is relatively easier to interpret. Plotting the estimated factor loadings for the given time points describes the shape of the non-linear relationship over time.

The loadings for the shape construct are termed *basis weights* because they reflect the basis by which the shape of the change can be interpreted and they provide the basis by which the change influences on the variances and covariances of the variables is evaluated (Duncan et al. 1999; McArdle & Nesselroade, 2002). Specifically, the mean of the shape factor reflects the expected amount of change, weighted by the estimated loading at each time point. The variance of the shape construct reflects the individual differences variability around the grand slope value. The variance of a variable at any given point in time is therefore decomposed into three sources: a part that is related to variability around the grand mean (i.e., intercept or level), a part that is related to variability around the grand slope (as determined by the basis weight for the time point in question), and a residual error component. Because a residual variance component is estimated for each variable at each time point the unreliability and uniqueness of each measured variable is accounted for and thus we can speak of these models as being *latent growth curve models* (McArdle & Nesselroade, 2002).

*Piece-wise Models.* In the previous examples, we used the passage of time as the index of change in the growth curve model. Another approach to conceptualizing time may be to extend the two-wave difference score models discussed previously to multiple waves. For the piece-wise model shown in the top panel of Figure 5, a factor is needed for every measurement
occasion. The first factor for this type of model represents the initial status rather than intercept as in previous growth curve models, but the factor loadings are still fixed at 1.0 for all occasions. The remaining factors represent differences in the scores of a measure between adjacent occasions (differences in scores between Time 1 and Time 2, between Time 2 and Time 3, and between Time 3 and Time 4). The weights for a factor representing differences in scores between occasion \( t \) and occasion \( t + 1 \) are fixed to zero for the first \( t \) occasions and -1 for the remaining occasions. Means and the variances of the first factor reflect means and variances of measures on the first occasion, while the means and variances of the remaining factors reflect means and variances of differences in scores between adjacent occasions.

Covariances between factors can also be estimated that reflect the extent to which initial and difference scores on adjacent occasions covary. When all covariances among factors are included in the model, the piece-wise model becomes saturated and fits the data perfectly. Accordingly, the residual variances are unnecessary and must be fixed to zero. Restrictions can be imposed on such a saturated piece-wise model, allowing for estimation of the residual variances and allowing degrees of freedom for model fit assessments. For example, the covariances between non-adjacent difference scores might be essentially zero. Accordingly, only covariances between adjacent factors might be estimated along with the residual variances allowing for two degrees of freedom with four measurement occasions.

**Spline models.** The simple polynomial, the level and shape, and the piece-wise models can be combined into a fourth class of model referred to as spline models. In a spline model, a focal point, sometimes referred to as a “knot”, is determined where the growth process is expected to change in a meaningful way. A spline model allows the researcher to model the growth process up to the knot in a different manner than growth after the knot. For instance, the middle panel of Figure 5 illustrates a spline model where rapid change is expected between the first 2 time points but a much slower rate of change is expected from Time 2 through Time 4. In this model, a restriction on the factors is imposed so that the slope would be constant between Time 1 and Time 2 (up to the knot at Time 2) and constant (linear) from Time 2 and Time 4 (after the knot located at Time 2) but at different rate. The first factor is an initial-status factor representing scores at Time 1, the second factor is a piece-wise factor that represents the difference between scores at Time 1 and Time 2, and the third factor represents a linear pattern of change from Time 2 through Time 4. In addition, a common residual variance and the covariances among all three factors are estimated. The second factor represents the difference score between Time 1 and Time 2 but also can be interpreted as the linear rate of change between those time points as well.

While the example presented reflects linear growth both prior to the knot and after the knot, any combination of the previous models could have been used as long as there are a sufficient number of time points to both support the model structure and preserve some degrees of freedom for model fit tests.

**The no-change or no-growth model.** A common baseline model that can be fit is one that assumes no change and simply estimates the variance around a grand, time-invariant mean level. In this model, there is only a single intercept or level factor as shown in the bottom panel of Figure 5. Models of the possible change relationships are then fit and compared to this baseline model to evaluate whether there is evidence for (significant) change in the time-ordered data. Evidence of change will emerge if the change process has affected any of the basic descriptive moments of the time-order variables (i.e., variances and covariances as well as mean-level differences; McArdle & Nesselroade, 2002).
Latent growth Curve Models with covariates. The models that have been discussed thus far can be considered as comprising the first of two stages. In this initial stage, we illustrated various approaches to conceptualizing the change or growth process within an individual over repeated measurements. The potential second stage of modeling that one might consider is to then predict the individual growth curve parameters in the same manner as one might attempt to predict a factor score in a confirmatory factor analysis model. As latent variables, the level and shape factors can easily be treated as independent or dependent variables in an expanded model. For instance, a variable such as “age” may be used to predict the initial status of a trait, and subsequent change in status over time may be a predictor of some other trait or behavior.

Multivariate Growth Curve Models. In the univariate growth curve modeling contexts previously described, a primary objective is to understand how changes over time occur for a single outcome measure. Univariate growth curve modeling can be extended to multivariate growth modeling, where the primary objective is to model univariate growth in the context of multiple parallel measures and relate those multiple outcomes to each other. Regardless of whether the goal of the analysis is a univariate model or a multivariate model, the initial stage is to investigate the models of change for all variables separately and treat these univariate models as building blocks for creating the more complex multivariate models.

The first type of multivariate model to be presented is often referred to as an associative LGM (Duncan et al, 1999). Like the previous discussion of LGMs with covariates, the univariate growth curve is determined in stage 1. An appropriate conceptualization of growth is determined for each developmental variable that is repeatedly observed over time. Once the univariate growth structure is determined, stage 2 entails modeling the covariances between all latent constructs in the model as shown in Panel A of Figure 6. This type of model is applicable when multiple constructs or processes are measured over time and the researcher is interested in how the latent aspects of one construct relates to the latent aspects of another construct. For instance, let us presume that the two constructs in Panel A of Figure 6 represent the traits of conscientiousness (1) and perfectionism (2). The associative LGM could be used to determine whether initial status in perfectionism is related to initial status in conscientiousness, whether initial status on one variable is related to the change that occurs over time in the other variable, and whether change in both constructs are related to each other. A simple extension is to replace the certain correlations with structural relationships and address whether initial status in perfectionism predicts change in conscientiousness or whether initial status in conscientiousness predicts change in perfectionism. In these models, the direction of the structural relationship is critical in that change over time cannot predict initial status.

A second type of multivariate LGM is the factor-of-curves model suggested by McArdle (1988) to determine whether higher-order factors can adequately represent the relationship between several univariate growth functions. In the factor-of-curves model shown in Panel B of Figure 6, a simple linear slope univariate growth function is developed for each three measures. The only difference between the linear slope functions at the lower level of the factor-of-curves model and the model presented in Figure 4 is the absence of the covariance between intercept and slope (which is now implied by the higher order factor) and the latent construct means (which are reproduced as a function of the higher-order factor means). Two higher-order latent factors are then included that represent the common intercept (Lc) and common linear slope (Sc). In this case, the higher-order latent variances are estimated and one factor loading per construct is fixed at a value of 1.0 for identification purposes; however, the variances could just have easily been constrained to 1.0 and all loadings freely estimated. It is often the case that additional
constraints, such as equating the loadings, are needed to fully identify the model parameters. Extending the example from the previous paragraph, let us presume that the third construct (3) is dutifulness. The factor-of-curves model may answer the question of whether conscientiousness, perfectionism, and dutifulness have a common initial status ($L_0$) and a common pattern of change over time ($S_0$).

The third type of multivariate LGM, the curve-of-factors LGM, can be used when modeling the growth function of a trait that has multiple indicators available at each measurement occasion. In Panel C of Figure 6, the latent factors $T_1 - T_4$ are measurement models for a trait with three indicators that has been assessed over four time points. The trait then is modeled as a higher-order LGM where the growth function is determined for the change in the construct over time rather than the change in the observed variables. For example, let us presume that the diagram in Panel C of Figure 6 represents three aspects of perfectionism (self-oriented [$V_1-V_4$], other-oriented [$V_5-V_8$], and socially prescribed [$V_9-V_{12}$]) measured using the Multidimensional Perfectionism Scale (MPS: Hewitt & Flett, 1991) on a sample of participants over four time periods. The curve-of-factors LGM allows us to determine the nature of change in the truly latent construct of perfectionism over time where the observation of perfectionism at each time point is itself a latent construct ($T_1 - T_4$) with multiple indicators.

The final type of multivariate model to be presented is an extension of the two-wave difference- and residual-score approaches discussed in section IIa of this chapter. Figure 7 presents a bivariate latent difference score model (LDS: McArdle & Hamagami, 2001; McArdle & Nesselroade, 1994). This type of model assumes that there is an underlying dynamic process that has been observed over a well-defined period of time. In the bivariate example in figure 7, there are two processes measured by single indicators over four time periods each ($h_0 - h_1$ & $g_0 - g_1$). In this model, changes in a process ($h$ and/or $g$) as a function time ($\Delta y[t]/\Delta t$) are both constant to a fixed slope and proportional to the previous measurement ($\gamma h[1]$). The $\alpha$ and $\beta$ parameters are the group or fixed coefficients, and variances on the latent variables $\Delta y_h$ and $\Delta g$ (represented in the model by $h_0^*, S_h^*, g_0^*, S_g^*$) allow for individual differences. The $\gamma$ parameters represent the dynamic coupling, or lead-lag relationship, between the two processes $h$ and $g$. For instance, the parameter $\gamma_h$ reflects the influence of process $h$ at time $t - 1$ on the change in process $g$ from time $t - 1$ to time $t (\Delta g)$ while the parameter $\gamma_g$ reflects the opposite influence. As in the previous growth curve examples, there are common residual error terms ($\theta^2_y$) and covariances ($\rho$) between the exogenous latent variables ($h_0^*, S_h^*, g_0^*, S_g^*$). The first three multivariate models described in this section consider change as a whole over the entire length of the time series of observations. The LDS model addresses the relationship between processes between successive time points. Returning to the initial example in this section involving just conscientiousness ($h$) and perfectionism ($g$), let us suppose that a researcher is interested in the role of conscientiousness in predicting change in perfectionism, but the researcher is interested at a more specific level than just “initial status predicts change”. Interpretation of the $\gamma_h$ parameter addresses whether conscientiousness serves as a leading indicator of perfectionism, while the $\gamma_g$ parameter reflects the opposing influence. That is, does change in conscientiousness from one observation to another serve as an antecedent event to change in perfectionism during the next successive interval?

d. Growth Mixture Distribution Modeling

A relatively new and increasingly popular branch of methodological approaches to longitudinal modeling seeks to investigate possible heterogeneity among individual patterns of growth (Muthén & Shedden, 1999; Nagin, 1999; Li, Duncan, & Duncan, 2001). Known
generally as Growth Mixture modeling, these techniques seek to identify clusters of similar
growth patterns across time. Since both the number of groups and their composition are
unknown, these models are somewhat exploratory in nature. Traditional growth curve models
that rely on structural equation modeling (and random-coefficients modeling) can account for
differences in the form of growth only for the sample as a whole or for subgroups if they are
known a priori. Growth Mixture modeling is useful to capture phenomenon that should have
distinct subgroups of individuals who are defined by their common patterns of intra-individual
change. For example, some individuals may never display any substance use, others may exhibit
moderate use that decreases with time, while a third group may display high use that continues
across time.

Growth Mixture modeling is a general class of technique that includes Latent Trajectory
Analysis (Nagin, 1999). A mixture model makes use of a mixture distribution, which is merely a
distribution of responses that is assumed to be comprised of more than one different distribution.
These component distributions may be of the same shape but differ only in location, or they can
vary in shape, or both. For example, the distribution of height among college students may be
thought of as a mixture of two normal distributions, one for male students and the second for
female students (see Rindskopf, 2004). In extreme cases, the unique parameters describing the
component distributions results in a mixture distribution that is multi-modal; however, such
obvious potential markers of a mixture distribution are not common (Bauer & Curran, 2004).
Growth mixture modeling has been developed for use with four unique distributions: (a) the
*normal* distribution used for multivariate normal data; (b) the *zero-inflated Poisson* used for low-
frequency count data; (c) *censored Normal* used for skewed and truncated data; and (d) *logit*
used for binary data (Nagin, 1999; Nagin & Tremblay, 2001). Applications to other forms of
distribution are possible but as yet have not been developed.

In a particular model, all groups are assumed to arise from the same distribution being only
distinguished by unique parameter estimates for each group. Like most other latent curve
approaches, Latent Trajectory Analysis uses a basic polynomial form for each mean group
trajectory. Thus, actual growth within a group may take the form of a linear, quadratic, or cubic
curve, for example. The model does allow testing for similarities across the group’s trajectories.
One very valuable attribute of the growth-mixture approach is that it allows examination of the
relationship between individual manifest variables and an estimate of the probability of group
membership. Although individuals are assigned to a latent group based on their highest
probability of group membership, the uncertainty of categorizing individuals into unobserved
groups is accounted for in estimating other parameters of the model (Nagin, 1999).

There are a few limitations in the Growth Mixture modeling approach. The first centers on
deciding on the optimal number of groups (Bauer & Curran, 2003a, 2003b; Muthén, 2003;
Rindskopf 2003). Typically, information criterion methods such as the Bayesian Information
Criterion (BIC) and the Akaike Information Criterion (AIC) have been used. However, these
methods do not yield formal statistical tests and only measure relative fit.

Second, it is possible that a single skewed distribution can be interpreted as a mixture of
distinct distributions (Bauer & Curran, 2003a). One limiting assumption of Nagin’s approach is
that it sets the variance of the group trajectories to zero, although other versions of the model
make the variance identical across groups (Muthén, 2001). This assumption means that any
individual differences apart from group membership are not accounted for by the model.

A similar technique known as Latent Transition Analysis (Collins, 1991; Collins & Wugalter,
1992) is an example of a latent Markov-Chain model (Meiser & Ohrt, 1996) and can be viewed
as an extension of Latent Class Analysis (LCA). Traditional factor analysis is theoretically similar to LCA. Whereas factor analysis relates several continuous observed variables to a set of continuous unobserved variables, Latent Class analysis uses information from several categorical observed variables to classify individuals into two or more unobserved groups. Because it uses information from multiple indicators, Latent Transition analysis takes measurement error into account.

Estimation of Latent Transition models is based on the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). Parameters of the model that can be estimated include: the proportion in a particular latent status at Time $t$, the probability of a response to an item given membership in a status, and the probability of membership in a status at Time $t+1$ given membership in a particular status at Time $t$. This final set of parameters forms the latent transition probability matrix that lists the probability of switching to a different status at each measurement occasion. These attributes make Latent Transition analysis well-suited for examining models of stage development. One key difference between Latent Transition analysis and Growth Mixture modeling is that in Latent Transition analysis the number of unobserved groups must be specified. In this context an unobserved group is referred to as a latent status.

At least three general criteria can be applied to verify the validity of the classes that these techniques indicate. The first criterion is replicability. Are the same classes or groups identified when the analyses is conducted a number of times, either using new samples or randomly dividing large samples into smaller samples (i.e., the basic cross-validation approach; see Cudeck & Henley, 2003)? The second criterion is interpretability; that is, do the growth patterns that typify each class make theoretical sense? The third criterion is predictability. Given the identified classes or groups, is there a set of unique predictions that can be derived about the behavior of each class or group on a set of variables that are independent of the variables used to identify the classes?

e. Time Series Analysis

The methods for the analysis of change that have been discussed thus far, including growth curve modeling and growth mixture distribution modeling, are generally implemented in designs involving a sample of participants and a smaller but adequate number of repeated observations where the number of participants is greater than the number of observations made on each participant. Increasingly, longitudinal research designs in the social sciences are resulting in analytic situations where individuals are measured repeatedly over a large number of intervals where 50 or more observations provide reasonable parameter estimates (Box & Pierce, 1970). Such time series designs can be considered as the ideal longitudinal design (Velicer & Fava, 2003). For instance, tracking mood states in hourly intervals can be made possible through technological advances such as the use of pagers, cell-phones, or hand-held computers, resulting in the ability to collect a long series of behavioral observations. The goal of a time series analysis is to identify time-related patterns in the sequence of numbers where the patterns are correlated, but offset in time. One could also assess the impact of one or more independent variables on the time series or perform forecasting. Generally, time series designs involve single subjects or a small number of subjects that are aggregated in some manner.

The need for a time series analysis arises when considering the appropriateness of classical methods such as multiple regression. The primary criteria for use of time series procedures rather than multiple regression analysis is the inherent dependency that results from making repeated observations of the same participant or group of participants, referred to as an autocorrelation. Analysis of time-series data in the presence of an autocorrelation using multiple regression
Methods for the Analysis

Techniques is an explicit violation of the assumption of independence of errors. As a result, Type I error rates would be substantially increased. In addition, “false” patterns may either obscure or spuriously enhance intervention effects unless the autocorrelation is accounted for in the model.

There are two primary classes of time series analysis differing by the domain in which they are applied. Time series in the time domain are conducted using the Box-Jenkins Auto-Regressive Integrated Moving Average (ARIMA: Box, Jenkins, & Reinsel, 1994) class of models where the pattern of change in a dependent variable is assessed over time. Analyses in the spectral domain, such as Fourier-Spectral-analyses, decompose a time series into its sine wave components. The ARIMA model has become the most common time series application in behavioral research (Velicer & Flava, 2003), providing a means of modeling the serial dependency in repeated measures data and thus provide valid statistical inference.

The basic properties of an ARIMA model are characterized by three parameters. The key elements of an ARIMA (p, d, q) time-series analysis are the lingering effects of preceding scores called auto-regressive elements (p), trends in the data called integrated elements (d), and the lingering effects of preceding shocks called the moving average element (q). All ARIMA models also have random process error terms called shocks, but differ in the order (how many preceding observations must be considered when determining the dependency) of the p, d, and q parameters. A time series analysis consists of three steps: (a) identifying which mathematical model best represents the data, focusing on the autocorrelation function, potential cyclic patterns, auto-regressive components (p), and moving-average components (q); b) reconfiguring the dependent observed variable into a serially independent variable through a transformation appropriate for the identified model; and (c) estimating the model parameters through generalized least squares and examining the residuals for unaccounted patterns.

Time series can be used to answer research questions involving autocorrelation patterns (“Are there linear or quadratic trends?” or “Do previous scores/shocks have an effect?”), cycles and trends (“Are longer trends separate from ‘local’ patterns?” or “Are there seasonal, periodic, or cyclic patterns over time?”), forecasting (“What is the predicted value of observations in the near future?”), or covariates (“Are there predictors?”). Time series can also be used to assess interventions. Is there an impact of an intervention/treatment, after accounting for patterns? Is the impact abrupt and permanent (e.g., man-made or natural disasters) or abrupt but temporary (e.g., New Year’s resolutions).

While time series analyses are typically performed on single-subject data, they can also be used to incorporate data from multiple participants or compare across individuals or groups to assess the degree of similarity in the patterns for different variables or populations. In pooled time series analysis (Hsiao, 1986; Dielman, 1989), all observations for all participants are included in a single vector, and a patterned transformation matrix is utilized to convert the serially dependent variable into a serially independent variable. Another alternative is meta-analysis, where individual participant time series are combined rather than individual studies. However, the meta-analytic approach is difficult in that there is a lack of statistical time series information in the published literature (many reports still rely on visual analysis) and an appropriate definition of effect size is needed for time series data.

Multilevel or mixed modeling, as was described in the previous sections, can be considered as a means of utilizing data from multiple participants, where the elements of the time series are nested within individuals, resulting in a 2-level hierarchy. A multilevel approach to time series analysis is more difficult in the SEM context due to the necessary length of the time series (> 50), but can easily be conducted using traditional multilevel software (for example, HLM or SAS PROC...
Multivariate time series models are then models where there are multiple measures at each time point for the same individual and each variable is a time series. A basic approach can be to determine the cross-lagged correlational structure between the multiple variables, where lag refers to the time relationship between two variables. If one variable can be conceptualized as a dependent variable and the remaining variables can be considered covariates, then a concomitant variable time series analysis (Glass, Willson, & Gottman, 1975) can be conducted as a direct analog to the analysis of covariance.

While the use of an SEM program to perform a time series analysis is not generally recommended (Velicer & Flava, 2003; cf. dynamic p-technique SEM, below), both univariate (single variable) and multivariate time series data can also be represented as special cases of SEM (van Buuren, 1997). A related approach is dynamic factor analysis, an extension of p-technique factor analysis (Cattell, 1952, 1963, 1988). While non-dynamic p-technique factor analysis has major concerns regarding the serial dependency inherent in the time series and can result in substantially under estimated factor loadings (i.e., due to positive autocorrelations; Wood & Brown, 1994), dynamic factor analysis permits the serial dependency (Hershberger, 1998). Nesselroade and Molenaar (1999) proposed combining pooled time series techniques with dynamic factor analysis to overcome the limitation of the number of observations in the time series needed for stable estimation of the population covariance matrices. The next section describes a further extension of dynamic factor analysis.

f. dynamic single-subject repeated-measures designs: p-technique

A powerful, yet relatively under utilized technique to model developmental phenomena is the dynamic single-subject repeated-measures design, or p-technique. Dynamic p-technique factor analysis has seen some application in the field of psychology in, for example, the domains of mood, personality, and locus of control (see Jones & Nesselroade, 1990, for review); however, the technique, as yet, has not been fully exploited as a potentially powerful research tool. Nesselroade and Molenaar (1999) have argued that the intensive study of a single individual allows one to model the truly dynamic interplay among variables. Such designs are ideally suited to examine questions regarding topics such as the person-fit debate (Fleeson et al., this volume) and the social-personality nexus. Dynamic p-technique SEM has the additional advantage of modeling change relationships as latent constructs (i.e., error free) with varying degrees of potential lagged influences. Although dynamic p-technique SEM models can be fit to a single individual, the broader usefulness of this approach emerges when one compares the resulting dynamic models of change across a sample of individuals. With this approach, the key sample-size issue is insuring sufficient data points for each individual to establish a well-conditioned model for each individual. The number of persons needed to make reasonably sound nomothetic generalizations is, therefore, relatively small.

The basic idea behind dynamic p-technique SEM is that a set of indicators measured repeatedly over time will yield a covariance structure that can be modeled using SEM (Cattell, 1963). Indicators of a construct will ebb and flow over time in a uniform manner such that they will reflect an underlying latent construct that is defined on the basis of the ebb and flow of change. The relationships among multiple constructs can then be assessed and compared on the basis of their cross-time changes with one another. Such a time-order data matrix easily captures contemporaneous dynamic processes, but does not capture lagged dynamic influences. To address this limitation, dynamic p-technique SEM utilizes the inherent time-ordered information of such a data matrix to create a lagged covariance matrix wherein the effects of the constructs at
time \( t \) can be evaluated for their impact on the latent constructs at time \( t + 1 \) (Hawley & Little, 2003). Figure 8 depicts a basic block Toeplitz covariance matrix that is modeled in a dynamic p-technique SEM.

The structure of such a lagged covariance matrix contains three distinct elements. The simultaneous or synchronous relations among the three variables are represented twice, at Lag 0 and again at Lag 1, in the triangles directly below the major diagonal. At each lag, the variances of the variables are located along the major diagonals and covariances are located off the diagonals. For the most part, the corresponding elements between these two sections would be nearly or exactly identical (see Hawley & Little, 2003). The lower quadrant of the lagged covariance matrix contains the lagged information among the variables, which reflects two sources of lag information. The first source of information is the auto-regressive lagged relations between each variable. This information is represented along the diagonal of this lower quadrant \( (AR_{1,1}^*, AR_{2,2}^*, \text{ and } AR_{3,3}^*) \), that is, a variable’s correlation with itself between lag 0 and lag 1. Cross-lagged relations among the variables are represented in the upper and lower triangles of this quadrant (e.g., \( CL_{1,2}^* \), \( CL_{1,3}^* \), and \( CL_{2,3}^* \)) which represent, for example, covariation between variable 1 at lag 0 \( (V_1) \) and variable 2 at lag 1 \( (V_2) \), \( V_1 \) and \( V_3^* \), \( V_2 \) and \( V_3^* \), and so on (see e.g., Molenaar, 1985; Molenaar, De Gooijer, & Schmitz, 1992; Wood & Brown, 1994).

Both time-series and dynamic p-technique SEM are particularly useful for intensive repeated measures designs that are geared to understand dynamic change processes. Like time-series analyses, key advantages of the dynamic p-technique include accounting for the auto-correlation among indicators as well as cross-lagged influences. Further advantages of dynamic p-technique include the inherent correction for measurement error (i.e., multiple indicators of latent variables are employed) and nearly limitless expandability of the model to incorporate static covariates, time-varying effects, and static outcomes (see survival analysis, below). With multiple-group capabilities of SEM programs, comparing models across groups of individuals allows nomothetic assessments of the similarities and differences in the dynamic patterns among individuals.

**g. Survival Analysis**

Often in longitudinal research the outcome variable of interest is the absence or presence of a specific event. In this context an event is defined as a qualitative change in state that occurs at a precise moment in time, including such occasions as incarceration, graduation, divorce, or death. Techniques to deal with such events fall under the rubric of survival analysis, and are primarily concerned with the length of time before the event occurs. The term survival analysis comes from biostatistics, but in other fields is also known as event history analysis, lifetime analysis, and reliability analysis (Allison, 1984). One unique circumstance with this type of data is when an event has not occurred for an individual by the end point of data collection. Such data is considered to be censored. Presumably, if the study had gone on indefinitely the event would have (or could have) taken place for the individual. Because of the censoring issue standard regression models predicting the time to an event will be biased. Remedial measures such as deleting or assigning the maximum time to the censored cases may only exacerbate the problem. Models are typically instantiated for events that are not repeatable, or the case in which only the first occurrence is of interest (for models of repeatable events, see, Blossfeld, Hammerle & Mayer, 1989; Klein & Moeschberger, 1997).

There are two basic functions that are central to survival analysis, the survivor function and the hazard function. The survivor function is the probability of survival through a certain time period, say \( t \). The survivor function for Time \( t \) is the probability that an individual lasts at least past Time \( t \), consequently the survivor rate can not increase. The hazard function, more
Methods for the Analysis

commonly known as the hazard rate, is directly related to the survivor function. For the special case in which events take place in discrete time periods the hazard rate is simply the probability that the event occurs for an individual at a specific time given the event had not yet occurred for the individual. For continuous models the basic idea remains the same but is not as mathematically simple.

Some events, such as graduation, can only occur at certain points in time. While other events can take place at any given instant, they are often only measured in discrete time periods, such that the researcher only knows the month or year of event occurrence. For data such as these a discrete time survival model is necessary (Allison, 1982; Willett & Singer, 1993; Yamaguchi 1991). A typical discrete time survival analysis utilizes a person-period data set in which each individual has a separate record for each time period. The event variable is dummy-coded and listed with each person-period record.

Historically, most survival analysis models treat the hazard rate as a continuous function of time (Allison, 1995). These continuous survival models can be classified into two types, parametric models such as the accelerated failure time model (Kalbfleisch & Prentice, 2002) and semiparametric models such as the proportional hazards model, sometimes referred to simply as a Cox model (Cox 1972). The key difference between the two models is in the function of time. Parametric models assume a known distribution for survival time while semiparametric models do not. The basic form of both discrete and continuous models is a regression equation predicting the log of the time until event occurrence. Accelerated failure time models are similar to censored normal models. Several other types of distributions may be specified for time including log-logistic and Gompertz and Weibull (Blossfeld & Rowher, 2002). In the semiparametric Cox model, for the case where none of the predictor variables vary with time, the ratio of the hazard for any individual compared to any other individual is a constant. This fixed ratio is termed the proportional hazards assumption. However, this assumption does not hold when dealing with time-varying covariates. Unfortunately, most semiparametric models are still referred to as a proportional hazards model even when the hazards no longer meet this assumption. Additional models that may be useful include models of competing risks in which the type of event must be distinguished (Klein & Moeschberger, 1997).

III. General Measurement and Design Issues

a. Appropriate Calibration of Time

A crucial design issue in the study of development is the choice of time-unit calibration to index change. At least five categories of time-unit calibration can be used. A first calibration point is birth. Indexing change as a function of age in years, months, or days is clearly the most common metric used to examine developmental changes. For most studies of the development of temperament and personality that focus on childhood and adolescence, this time-unit is likely a sufficiently sensitive index of time. However, particularly during middle-age, years since birth may not be an appropriate index of life age (see Helson, Soto, & Cate, this volume).

A second calibration point is death. Time from death can be a very powerful index of change for many developmental phenomena, particularly those related to physical and mental functioning. For example, the steep decline in functioning approximately five years prior to death (i.e., terminal drop) can only be adequately indexed as time from death.

A third calibration point, which may be termed episodic time, is the time before and after a particular event such as puberty, marriage, first-job, first-born child, last promotion, retirement, and so on. Such normative life-events are particularly relevant for developmentalists who wish to understand the developmental impact of these important milestones along the life course. Critical
Methods for the Analysis

developmental questions related to the onset (early, on-time, late) of these milestones can be addressed directly using this form of calibration.

A fourth calibration point is experiential time. Experiential time is related to episodic time, but the focus is strictly forward in time. Here, change is indexed by the amount of time that a particular individual has experienced a particular status, influence, or context-related impact. Grade in school is a classic example of experiential time. A group of children who have experienced only two full years of elementary curriculum could be expected to be developmentally different than a group of children matched on the same age but who have experienced two and half or more years of elementary curriculum.

A fifth calibration metric is biological time. McArdle and Nesselroade (2002) provide a nice example of biological time calibration based on the peak of the growth spurt across a set of individuals. Related to event time (i.e., the peak of growth), this approach to calibration allows one to examine the degree to which the shape of the growth pattern is invariant across individuals even though the intercept or onset of the growth curve can vary across individuals.

An important consideration is that these different calibration points are not exclusive. For example a study of the effects of marriage could use the age since birth at which a person was married as well as the number of years that a person has been married. In one set of analyses, experiential time (years of marriage) could be used to index developmental changes but chronological age could be used as a covariate, mediator, or moderator of the developmental growth functions. Similarly, for other developmental questions, chronological age could be used to index developmental change while using experiential time as a covariate, mediator, or moderator.

b. Appropriate Interval of Time

In addition to the issue of calibrating time is the selection of time interval and the time duration needed to accurately capture the developmental process of interest. The issue of time interval relates to the frequency and distance between measurement occasions. For some developmental processes, the expected rate of change may be quite rapid, while for others the process may be quite slow in its progression. Too often, however, developmental processes are measured at intervals of convenience rather than intervals of sufficiency. For example, many developmental studies of childhood and adolescence conduct yearly assessments. Such a time interval would clearly be insufficient to capture inherent differential effects related to changes within a school year as opposed to changes that occur between school years (i.e., school effects vs. summer effects).

In terms of time duration, a proper longitudinal design must assess individuals over a sufficient time span such that the expected change period is captured. A two or three-year study of temperament in infancy and toddlerhood, for example, may be quite sufficient to track a non-linear unfolding of such phenomena. On the other hand, the period of time needed to capture change in higher-order personality constructs such as agreeableness in adulthood may require a decades-long study. In other words, because most developmental phenomena are globally non-linear, in order to have the fidelity to capture the curvature, the data collection period must begin well before and extend well beyond the expected bend points. Otherwise, more often than not, the data will appear to be well represented as a locally linear trend.

c. Appropriate Scaling of the Construct

Three basic problems with simple change scores were briefly discussed in section IIc of this chapter. The third basic problem, the lack of consistency in interpretability of the change score at different points in the distribution, warrants additional discussion here. Unless true interval level
of measurement is attained, the meaning of change depends on the initial score level. That is, a small change from an extreme score may imply a different degree of change than the same small change from a moderate score. This problem can be solved or at least alleviated by using another latent-variable approach called item-response theory (IRT) to obtain justifiable interval scale properties. The text by Lord and Novick (1968) is the classic resource, while Embretson and Reise (2000) and Hambleton, Swaminathan, and Rogers (1991) have contributed more recent texts.

Suppose, for example, that two participants are given a 25-item stress checklist with successive items increasing in degree of stressfulness. Upon analysis, the researcher finds that both participants indicated that they had experienced 15 out of the 25 stressors. Now suppose that participant “A” indicated experiencing the first 15 stressors, and participant “B” reported experiencing the last 15 stressors, thus participant “A” experienced the least stressful events while participant “B” experienced the most stressful events. Traditionally, scale scores are a combination of items checked or a sum of responses on a Likert scale, so in this example, both participants would be equally stressed. Under item response theory, however, the participants would receive different stress scores, with participant “B” experiencing a higher degree of stress than participant “A”. Under IRT an individual’s trait level is inferred from the individual’s responses on a measurable instrument by considering the characteristics of the responses themselves rather than solely determined by the number of items responded to. Thus, in this example, to assess a participants stress levels, IRT scoring weights the experience of the stressor with the stressfulness of the event, such that more stressful events are potentially “worth more”.

One of the defining characteristics of IRT is the property of invariance – that scale items and the sample of respondents are independent of each other (Hambleton et al, 1991). Unbiased estimates of item properties may be obtained from unrepresentative samples, while under traditional methods, an item that is determined to be low on a trait for a sample of high-trait individuals would appear to be higher on the trait for a low-trait sample. That is, under traditional methods, relatively low-stress event experienced by a sample of very high-stress individuals will appear to be more high-stress than it would when experienced by a low-stress sample. In such cases, it is highly likely that a greater proportion of the high-stress individuals have experienced the low-stress event while fewer of the low-stress individuals will have experienced the low-stress event. Under IRT, the stress level of the item would be unbiased and invariant (generalizable) whether it was administered to high-stress individuals or low-stress individuals. In fact, IRT allows items with the same properties and mean trait levels to be administered to any population by use of Rasch model scaling and by placing items and persons on a common scale (Embretson & Reise, 2000). For an explanation of the differences in discrimination between classical test theory and IRT, see Embretson and Reise (2000).

Under traditional methods, meaningful change scores can only be compared when initial score levels are equivalent (Embreton & Reise, 2000). In other words, if two participants have different initial scores, any change score representing the difference in performance between two test conditions is meaningless. A small deviation from a high initial score does not mean the same thing as a small score change from an average score unless a true interval scale level of measurement is achieved. If interval scaling is achieved through transformations, then it is still specific to that particular test administration. However, in IRT, change scores can be meaningfully compared even when the initial scores are unequal. This comparability is largely due to the interval scale nature of item trait-level parameters and individual trait-level parameters.
Bereiter (1963) indicated three basic problems with using a simple classical test theory difference score to indicate change as previously described. A fourth problem, not indicated by Bereiter, is whether the change score actually reflects change due to a condition or is simple error (Embretson, 1998). This issue relates to the problems with standard errors inherent in classical testing. Assuming that the standard error applies to all individuals and that a test is focused toward the “average” individual, then the task does not provide a basis to assess individuals at the extremes in the trait because there will be fewer items at their trait level. In addition, the standard error also depends on the specific population being tested, so comparisons across populations are inaccurate.

A special Rasch-family model, the multidimensional Rasch model for learning and change (MRMLC; Embretson, 1991) addresses the four difficulties of classical test theory by resolving the scaling and reliability problems found with standard “change” scores and removing some of the confounds that occur with initial status. Two of the problems are addressed by IRT in general. First, the Rasch model achieves interval scale properties (see Andrich, 1985; Fischer, 1995), and as an interval scale model, it is distribution-free. Second, the MRMLC, as an IRT model, provides individual standard error of measurement estimates. The MRMLC specifically addresses two of Bereiter’s (1963) dilemmas with change scores. The issue of paradoxical reliabilities is addressed by modeling individual change directly in a model that explains changing test correlations. Bereiter’s problem of the correlation between the initial score and the change score is also resolved by achieving interval scale properties (Embretson, 1998).

In summary, although there are several problems that have been observed with classical test theory approaches to static measurement and dynamic change, item response theory is able to adequately address such issues. This is done by achieving interval scale properties and by making items, tests, and trait measures platform-independent and population-independent.

d. Calibrating Developmentally Appropriate Measures

In a longitudinal study, it is sometimes necessary to use different measures of the same trait depending on the developmental stage of the participants. For instance, an attitude measure given to a toddler or pre-school aged child must necessarily differ from a measure of the same construct given to an adolescent or young adult, but a study of the development of such a construct over a broad age range may necessitate the use of multiple age-appropriate measures. Likewise, studies across ethnicities, languages, or cultures may also require some ability to meaningfully combine multiple group-appropriate measures of the same construct. Assuming the construct validity of the multiple measures is assured, it becomes necessary to equate or calibrate the different measures. Equating requires that each measure have the same reliability and tap the same construct, which is often unrealistic in practice. Calibration, including vertical equating techniques, requires the same construct to be tapped but reliability and scale information may differ, and is a weaker form of equating. This discussion will focus on two common IRT-based methods of linking utilizing a vertical equating approach.

In common item equating, items are shared across multiple forms. For example, two forms are created -- one appropriate for the younger portion of the age range of interest and the other appropriate for the older age range. Both forms would then include a subset of common items called linking items that would be appropriate for both age ranges. Responses to the linking items form the basis for equating the two forms and placing all responses across the age range on a comparable metric. In common person equating, respondents are shared across test forms. That is, a random subsample of study participants would be given both forms, where the forms do not
include linking items, and the subsample’s responses to both forms provide the basis for comparability. For additional explanations of item equating, see Wright and Stone (1979), Mislevy (1992), Linn (1993), Kolen and Brennan (1995), or Thissen and Wainer (2001).

References with a * do not yet appear in the text.


---

**Figure 1: Latent difference and residual scores.** Note: The variances of the exogenous factors $T_1$ and $\Delta$ are fixed at 1. $T_1$ and $T_2$ are measured with multiple indicators that are invariant over Time.

**Figure 2: Latent difference and residual scores.** Note: The variances of the exogenous factor $T_1$ is fixed at 1. In these simplified drawings, the disturbance factor reflects the latent change variance, $\Delta$. 

A) Latent Difference Score

B) Latent Residual Score
Figure 3. Traditional SEM Model of Individual Differences Change Relationships. Note. The variances of the exogenous factors at T1 are fixed at 1 to set the scale and identify the constructs. T1, T2, and T3 are measured with multiple indicators that are invariant over time and each corresponding residual is allowed to correlate to account for the potential dependencies in the unique factors.

Figure 4. Various forms of Latent Growth Curve Models. Note.
Methods for the Analysis

Figure 5. Various forms of Latent Growth Curve Models. Note.

A) Piece-wise Model

B) Spline Model

Figure 6. Various forms of Multivariate Latent Growth Curve Models. Note.

A) Associative LGM

B) Factor-of-Curves LGM

C) Curve-of-Factors LGM
Figure 7. A bivariate Latent Difference Score (LDS) model (McArdle & Hamagami, 2001).

Figure 8. Need a Figure Heading Here.