CHAPTER 23

Foundational Issues in the Contemporary Modeling of Longitudinal Trajectories

JOHN J. McARDLE

Defining and Describing Trajectories Over Time

Figure 23.1 is a plot of longitudinal data from one of my favorite studies—the growth of verbal cognition in a set of children, originally carried out in a study described by Osborne and Suddick (1972). In this study, a special set of children ($N = 204$) were measured in school on the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1970) on four separate occasions.

Figure 23.1a is a collection of dots and lines now commonly referred to as "longitudinal trajectories." The long history of this term is not widely known (see McArdle, 1988, 2001; McArdle & Nesselroade, 2003), but it is definitely popular (e.g., Raudenbush, 2001). One common definition is that "a trajectory is the path a moving object follows through space as a function of time" (*http://en.wikipedia.org/wiki/Trajectory*). In this case the people are thought of as moving through time (the X-axis), and the score they obtain on the WISC measure (Y-axis) is plotted as a dot; these dots are supposedly comparable over these ages. That is, we assume that the WISC score means the same thing at each time, so any score changes reflect changes in the people. Then the dots for the same person are connected by a straight line to form the individual trajectory. Of course, the scores could fluctuate in many ways between the dots (Wohlwill, 1973).

Figure 23.1b is a plot of the same WISC data, but here the information is presented in a scatterplot matrix. This shows the within-time univariate distributions on the diagonal and the time-to-time pairwise relationships among all persons on the off-diagonals. In this type of plot we can envisage data from all the people, and this even makes it easy to spot a few odd scores (the dots away from the main plot). There are many ways to plot longitudinal data (e.g., see Wainer & Spence, 2004), but it is hard to make the connections achieved by

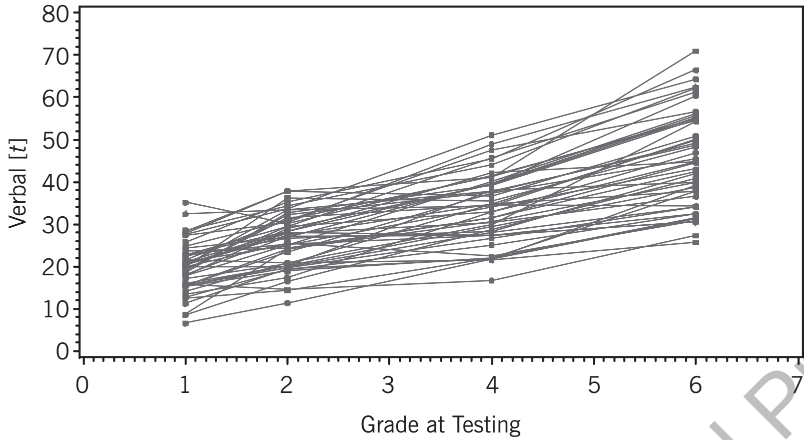

FIGURE 23.1a. Individual trajectories on the WISC Verbal scales ($N = 50$).

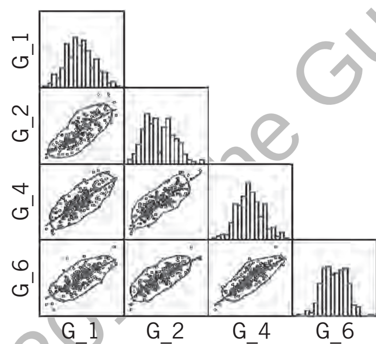

FIGURE 23.1b. A scatterplot matrix for four occasions of WISC data ($N = 204$). G, grade.

the simple lines and dots of Figure 23.1. In fact, it seems a bit easier to use Figure 23.1 to identify where a person has been and where he or she is going. Thus one advantage of the longitudinal trajectory plots is that they can help us tell a clear story about the person.

It has been my pleasure to deal with these longitudinal WISC data for about 20 years now, and I have tried many different kinds of analyses, including the beginnings of latent curve analysis (McArdle & Epstein, 1987), multivariate latent curve analysis (McArdle, 1988), multiple group factor analysis of latent curves (McArdle, 1989), autoregression versus latent curves (McArdle & Aber, 1990), multivariate analysis of variance (MANOVA) and the beginnings of latent change score analysis (McArdle & Nesselroade, 1994), and multivariate latent change score analysis (McArdle, 2001). I do not think I have found the best or most proper way to analyze these data, but I have certainly tried a lot of different forms of data analysis with these WISC data. I summarize my efforts here.

In a chapter on this topic Baltes and Nesselroade (1979) make a claim that, in studies of developing organisms, we often rely on some combination of information about (1) between-person differences, together with (2) within-person differences. In the prior literature on human development, these complex data collections are generally used for the study of intraindividual and interindividual differences. Instead of defining this in terms of algebraic relevance, they simply presented the five goals of longitudinal data analysis listed in Table 23.1.

What I am going to do now is use these five goals as my organization of our possible data analysis techniques and strategies. I am sure I will take some liberties with my reexpression of these earlier ideas, but I try to follow Baltes and Nesselroade (1979). Several decades ago longitudinal analyses were based largely on principles derived from linear growth models and formalized in terms of *analysis of variance* (ANOVA) techniques (e.g., Bock, 1975). In earlier work, these concepts were extended in the creation of randomized blocks and Latin squares (see Winer, 1962). Nowadays in research on human aging we often separate cross-sectional "age differences" from longitudinal "age changes" (see McArdle, 2009). Most of the newest analyses of developmental data analysis use information gathered from both cross-sectional and longitudinal selections in what are often termed "panel studies" (Hsiao, 2003).

The recent presentations on longitudinal data analysis are based on statistical procedures that combine these seemingly separate estimates. One way to do this is to create an "expected trajectory over time," in which the expected values are *maximum likelihood estimates* (MLE) and formal tests of hypotheses are encouraged (e.g., Hsiao, 2003; Miyazaki & Raudenbush, 2000; cf. McArdle & Bell, 2000). New computer programs for what are termed "latent curve" or "mixed effects" modeling allow model parameters to be estimated and used in making predictions and inferences. But this chapter is based on earlier work on longitudinal structural equation models (SEM; see McArdle, 2009). It is most important to recognize that we need to have some ways to understand these longitudinal data, and several theoretical statements have been formalized into "models" with useful statistical properties. These models allow us to consider alternative substantive ideas to fit these alternatives to our data and hopefully to make an informed choice about which alternatives are most useful. The great failure of any kind of analysis comes when a researcher reflexively applies any methods without thinking about what they mean and/or when the model predictions fail to look like the data.

I should state up front that, in my view, the current SEM approach is far less revolutionary than the past work. I think the new SEM techniques are not really much better than the old ones (cf. Raykov & Marcoulides, 2008). In my view, the classical techniques still form basic guideposts, and there are many worthwhile aspects of the older classical calculations. I mainly use SEM because it is convenient now.

TABLE 23.1. Five Objectives of Longitudinal Research

1. Direct identification of intraindividual changes.
2. Direct identification of interindividual differences in intraindividual changes.
3. Analysis of interrelationships in change.
4. Analysis of causes (determinants) of intraindividual change.
5. Analysis of causes (determinants) of interindividual differences in intraindividual change.

Note. Based on Baltes and Nesselroade (1979).

Issue 1: Studying Interindividual Changes with SEM

The very first issue we consider is how to organize information about the changes within a person. Baltes and Nesselroade (1979) used the term *intraindividual differences* but I now think it is better to state this as *within-person changes* (McArdle, 2009). There are two reasons for changing the prior terminology: (1) The term *within* is traditional ANOVA terminology to designate the separation of scores for the same person. This is in contrast to components that are "between persons." (2) The term *differences* rather than *changes* can then be reserved for models that imply the separation of different people.

A first important set of models for repeated-measures data has emerged from the "time-series" perspective and applied to panel data (see Andersen, 1971; Browne & Nesselroade, 2005). These models essentially suggest that the "future is predictable from the past" and that we should use this as a main feature of our analysis. The path diagrams of Figure 23.2 describe some options for these kinds of time series analyses of the WISC data. In general, a model is written in which the future deviation at time point [*t*] was predicted from deviation at the prior time point ([*t* – 1]) using a linear regression model with fixed parameters (b_0, b_1) and independent errors (*e*[*t*]). The utility of time series regression appeals to some researchers more than others (McArdle & Epstein, 1987).

The path diagrams of Figure 23.2 provide an exact translation of the necessary matrix formulation of these models (see McArdle, 1986, 1988), so none are provided. These path diagrams can be conceptually useful devices for understanding the basic modeling concepts. They are also practically useful because they can be used to represent the input and output of any computer program. In these path diagrams the observed variables are drawn as squares, and the unobserved variables are drawn as circles. Model parameters representing "fixed" or "group" coefficients are drawn as one-headed arrows, and "random" or "individual" characteristics are drawn as two-headed arrows. Here, the initial scores are random variables with no means but "random" variances ($\sigma_{[1]}{}^2$), and the effect of these scores on later scores is designated with regressions ($\beta[t]$). Standard deviations of the residual terms (Ψ_e) are also included (as in McArdle & Hamagami, 1991).

Figure 23.2a is a simple diagram in which everything in the past is used to produce (or predict) the future behaviors on the WISC. That is, the 1st-grade scores are used to produce the scores at the second grade, the fourth grade, and the sixth grade with unknown coefficients of proportionality ($\alpha[t]$). This implies that deviations from the mean for any person in the scores at later grades can be accounted for to some unknown degree by the early

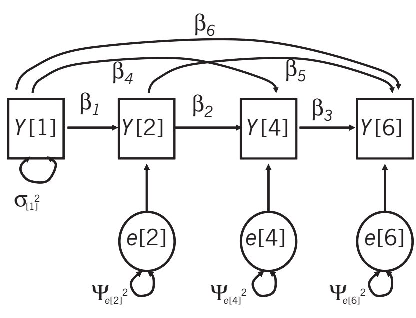

FIGURE 23.2a. A "fully recursive" Markov chain model with time-based effects.

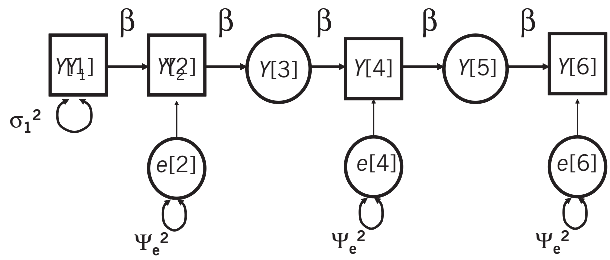

FIGURE 23.2b. An equal-time, equal-parameter "Markov simplex" model.

deviations from the mean. In a similar fashion, the second-grade scores are used to produce the fourth- and sixth-grade scores, and the fourth-grade scores are used to produce the sixth-grade scores. This kind of model is termed *fully recursive* because it requires as many parameters to be calculated as there are correlations observed in the data (Fig. 23.1b), so it is really not testing any major substantive idea here. To be clear, in this popular formulation, we are not really able to test whether or not the past predicts the future, or even vice versa, but we can use this approach to calculate the predicted values.

Figure 23.2a can have some important constraints. In a new version of this model, we can say that the only predictors needed for future scores are the ones in the immediate past. That is, the first affects the second, the second affects the fourth, and the fourth affects the sixth, and that is all that is needed to predict the future values. Because the time points are not equally spaced, we typically do not add further constraints to this model. This means we can eliminate three parameters from Figure 23.2a, and we can test this comparison of parameters as a formal hypothesis with three *degrees of freedom* (*df*). It is useful to restate that we are actually testing the importance of parameters that are not present, not the ones that are present. By fitting this alternative model, we can examine whether or not these simple time series restrictions seem reasonable for these WISC data.

We can add another important component to this model—a *latent variable* (LV), typically drawn as a circle with an arrow pointing directly to the square. This kind of LV is used because we want to assume that the time series process does not apply to the observed or *manifest variable* (MV), but the process applies to one key source of the variation in the MV—namely, the LV over time. Here we assume that there is one LV at each time, that the part of the MV that is not attributable to the LV is "unique," it may be partly "error of measurement," and it has the same variation at all times; this is exactly like a common factor model, but here the common factor has only one indicator, so it is not broad in any sense, nor is it tested in a rigorous way. But we should not underestimate the importance of this LV idea. One benefit of using LVs here is that we can estimate the process in the presence of random error (see Heise, 1975; Jöreskog & Sorbom, 1979). In this formulation, the time series for the LV is error-free, so we should find it to have predictions that are absent of error, and thus higher in accuracy. The model may fit the data better using this LV approach, but we do not have nor will we obtain estimated scores for the LVs. What we can examine is the effect of measurement error on our understanding of the process.

Figure 23.2b is an alternative and possibly simpler way to use LVs. Here we start with Figure 23.2a, but we then eliminate the parameters described earlier, and we add two LVs at the third and fifth grades to deal with the "unbalanced" data over time. In this conception, we assume that the children have verbal ability scores at each occasion, but we recognize that we did not measure them at these occasions. The main benefit of this type of LV model is that now we can consider a process that is measured over fairly equal units of time. That is, the interval of time between each LV is now 1 year, so we can constrain the parameter to be identical over time ($\beta[t] = \beta$). This simplifies the analysis because now we only have one parameter for any 1-year interval of time, and we have two additional *df*s to test the validity of this proportional prediction. Due to its simplicity, the prediction system is formally known as a "Markov simplex" model, but this seems to be misunderstood in the current literature (see Jöreskog & Sorbom, 1979; McArdle & Aber, 1990).

All of these time series models of Figure 23.2 were fitted to the WISC data, but they did not seem to fit very well; that is, the misfit index (χ^2) was large relative to the *df* (for details, see McArdle & Aber, 1990; McArdle & Epstein, 1987). The great success of time series models is that they allow us to make predictions about the future, and these predictions are fairly simple, but the failure to fit these models here means that these predictions were not very close to what had actually happened in the WISC scores. Part of this was the prior problem of "unbalanced" sequences (i.e., a 1-year gap, then a 2-year gap), but this seemed manageable (i.e., Figure 23.2b). But another part of the overall misfit of these models to the WISC data was clear—these time series models typically make an assumption of "stationarity"—that the variances (and covariances) remain the same over all times because the system has reached a "steady state of equilibrium" (see Heise, 1975). Equilibrium does not seem to be a reasonable assumption for verbal abilities in this young age range, and it is probably not a reasonable assumption for many other situations, either. Another assumption that was not tested is that the mean changes are not central to our understanding of the process. This problem does not create more of a misfit, but it can be a problem when the change for the whole group is a key part of the substantive questions.

To deal with some of these problems, a seemingly completely different set of models was also used. They are termed *latent curve models* (LCM), and they can be fitted by either approximate ANOVA methods or exact SEM methods (Meredith & Tisak, 1990; see McArdle, 1986, 1988; McArdle & Epstein, 1987; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002). This LCM approach typically starts by assuming we have a *trajectory* equation for each occasion for each person formed as the sum of (1) unobserved or *latent scores* representing the individual's initial level (f_0), (2) unobserved or latent scores representing the individual *change over time* or *slope* (f_1), and (3) unobserved and independent unique features of measurements ($u[t]$). In this model the arrows from the latent slopes to observed variables are a set of group coefficients or *basis weights* that define the timing or *shape of the trajectory over time* (e.g., $\alpha[t] = t - 1$). We create this carefully selected set of basis weights (as $\alpha[t] = [0,1,3,5]$) so it takes care of the unbalanced time delay between occasions.

One version of this type of model is presented as a standard path diagram in Figure 23.3a. As before, the path diagrams of Figure 23.3 provide an exact translation of the necessary matrix formulation (see McArdle, 1986, 2007). In this case the initial level and slopes are often assumed to be random variables with "fixed" means (μ_0, μ_1) but "random" variances (σ_0^2, σ_1^2) and correlations (ρ_{01}). The standard deviations (σ_j) are drawn to permit the direct representation of the correlations (see McArdle & Hamagami, 1991). One new feature here is the use of a triangle to represent a measured variable that has no variation—that is, the unit constant. The key reason this new option is added to the path diagram is to allow us to make hypotheses about the group means. In LCM we want to make hypotheses

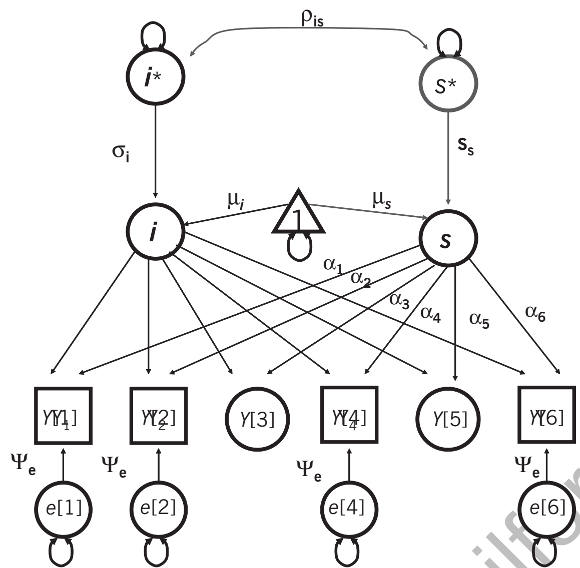

FIGURE 23.3a. A path diagram for a latent curve model with linear "slopes" based on weights $A[t]$.

about the group means and covariances that use the same parameters of proportionality (i.e., the basis $\alpha[t]$).

This LCM path diagram can also be interpreted as a two-common-factor model with means. The first latent factor score is an intercept or level score (labeled f_0), and the second latent factor score is a slope or change score (labeled f_1). The relationships between the latent levels f_0 and all observed scores $Y[t]$ are fixed at a value of 1. In contrast, the relationships between the latent slopes f_1 and all observed scores $Y[t]$ are assigned a value based on the time parameter $\alpha[t]$, which, depending on the application, may be fixed or estimated.

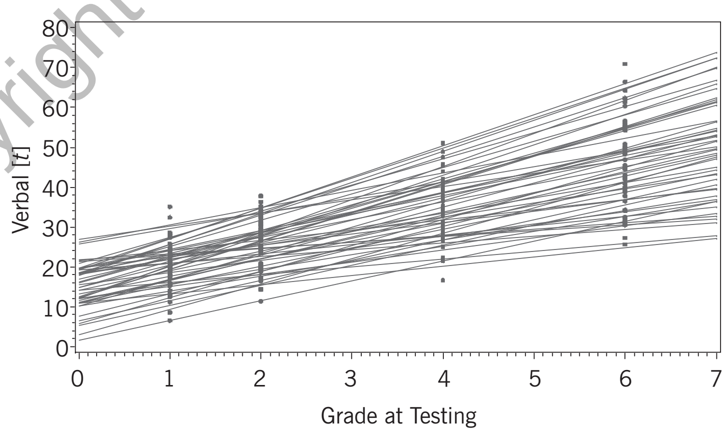

FIGURE 23.3b. Linear regression estimates of individual growth curves ($n = 50$).

For simplicity, the unique components ($u[t]$) are constant (Ψ_e) and uncorrelated with other components. In this longitudinal model the change score (f_1) is assumed to be constant *within* an individual, but it is not assumed to be the same *between* individuals. The latent variables are written in lower case (f_0, f_1) because they are similar to the predicted scores in a standard regression equation; that is, we do not use Greek notation because these scores do not need to be estimated (but see Figure 23.3b). Perhaps it is obvious, but this LCM is not the same as the prior time series model conception.

This simple linear LCM was fitted to the WISC data, and the results yielded a much better fit than found with other models (see McArdle & Aber, 1990). Several key issues can be seen in the simpler ordinary least squares (OLS) estimates of individual linear growth in Figure 23.3b. This plot is the result of extracting the longitudinal data for each individual and running a separate regression analysis for each person, whereas the WISC score was a simple linear function of time, and the plot included expected values. This yields estimates of linear equations with intercepts and slopes for each person, and this is a good way to see the key features of the LCM. First, the people vary at the first occasion, and this is a clear estimate of the intercept term in the LCM. In this figure it is clear that all persons are not alike—(1) we have an average intercept and a variation around this average; (2) we have a mean slope, a variation around this mean, and a possible covariance with the intercept score; and (3) we can see that the model does not fit all the data points (dots) in the plot. Thus we take the best parts of the LV approach, including the removal of independent unique variability, and the evaluation of systematic levels and slopes, and we add this to the individual and overall statistical indicators of goodness of fit.

Another interesting point is that the standard LCM is identical to the typical univariate repeated-measures ANOVA when the slopes are not allowed to vary (see McCall & Appelbaum, 1973). That is, the LCM approach offers a direct way to test the standard ANOVA as a simplified hypothesis (i.e., restricting the variance and covariance of the slopes to be zero results in "compound symmetry"). Although this standard ANOVA model may be reasonable in a number of circumstances, it is not reasonable with these WISC data because children do seem to vary in their slopes.

Although a linear scaling of the basis is very popular (see Singer & Willett, 2003), it is only one of many that could be used. For example, it is possible to add a lot of nonlinear complexity to the simple growth curve models for the study of within-person changes. For example, Wishart (1938) introduced a fundamental way to examine a nonlinear shape—the use of power polynomials to better fit the curvature apparent in growth data. The individual growth curve (consisting of $t = 1, T$ occasions) is summarized into a small set of linear orthogonal polynomial coefficients based on a fixed power series of time ($\alpha[t]$, $\frac{1}{2}\alpha[t]^2$, $\frac{1}{3}\alpha[t]^3, \cdots 1/p\ \alpha[t]^p$) describing the general nonlinear shape of the growth curve. A second-order (quadratic) polynomial growth model implies that the loadings of the second component are fixed to be a function of the first components (i.e., the derivative is linear with time).

The quadratic form of this basic model can be depicted as a path diagram as well (not included here). This can appear to be a bit complicated because it requires a third latent component with a basis that is related to the first one (i.e., $\frac{1}{2}\alpha[t]^2$), but all of this is done so that the implied change is linear with time (i.e., we add acceleration). Additional variance and covariance terms can be used to account for individual differences in these new LVs. In order to clarify these expressions, OLS quadratic regressions can be estimated for all the WISC scores on an individual basis (as in Figure 23.3b). Typically we find that introducing some curvature allows the model to approximate the data points more closely. Of course, a model of growth data might require this form of a second-order (quadratic), third-order (cubic), or even higher order polynomial model fitted to the data, and this family kind of nonlinear models are very popular (e.g., Bryk & Raudenbush, 1992).

One reason researchers like the LCM logic is that it permits many alternative extensions. A different alternative to the linear growth model was brought to light by Meredith and Tisak (1990)—the model proposed by Rao and Tucker in the form of summations of "latent curves." These innovative techniques were important because this made it possible to represent a wide range of alternative growth and change models by adding the benefits of the SEM techniques (McArdle, 1986, 1997; McArdle & Anderson, 1990; McArdle & Epstein, 1987; McArdle & Hamagami, 1991). Our use of this latent curve concept can be a minor adjustment to the LCM of Figure 23.3a: We allow the curve basis to take on a form based on the empirical data. In this example we simply write a model for the same person at multiple occasions in which the last two basis coefficients $a[3]$ and $a[4]$ are *free to be estimated* (the first two are still fixed at $a[1] = 0$ and $a[2] = 1$). The actual time of the measurement is known, but the basis parameters are allowed to be freely estimated so we can end up with different distances between time points—that is, an optimal shape for the whole curve. The estimated basis has been termed a "metameter" or "latent time" scale that can be plotted against the actual age curve for clearer interpretation (McArdle & Epstein, 1987). There are many ways to estimate these model parameters in an LCM framework, but this is very difficult to do with OLS, so these values are not estimated here.

The time series model and LCMs are only two approaches to the analysis of within-person changes that, while representing major modeling approaches, barely touch the surface of all other possibilities (e.g., see Walls & Schafer, 2006). These illustrations mainly show that any SEM, based on autoregressive (AR)-type models or LCMs, is used to make a prediction about the longitudinal trajectory based on a substantive issue. Thus both kinds of models have their place in the array of possible analytic frameworks, and choices between models are not often so easy. Some of these choices are simply defined by the data. For example, we might have a data set in which the individuals are measured repeatedly at more time points (e.g., $T = 100$). In this sense, this WISC collection of complete cases at a few time points ($T = 4$) is fairly easy to consider, even if the timing is unbalanced, and this allows us to go a bit further here.

Issue 2: Studying Interindividual Differences in Individual Changes with SEM

The second aspect of Baltes and Nesselroade's article (1979; see also Table 23.1) is the suggestion that we consider analyses focused on differences "between groups of persons in the way people change." In the WISC data we found that the original authors measured mothers' educational attainment levels before the children went to school at all; this showed a wide range across children, and we used this as the between-group variable (MOED). In one simple approach, we can just subdivide the WISC data into groups based on MOED and see whether or not the trajectories appear to differ.

At this time, we can consider four ways that the available SEM can be used to achieve these goals. We can (1) add measured variables (X) that define group differences into the previous expressions, (2) use the group variable to split the observed data into subsets for separate analyses, (3) define the grouping variable based on patterns of available data to deal with incomplete pieces, and/or (4) create latent groupings of people based on the pattern of available data. Of course, in any of these between-persons models the overall misfit can also be decomposed into individual components and outliers can be examined (e.g., McArdle, 1997).

Let us first add a measured variable (X = MOED) to characterize group differences directly into the previous linear expressions. Indeed, the most popular model used nowadays is the inclusion of the between-persons predictor in an LCM framework and having this static (one time of measurement) group variable account for systematic variation in the latent levels and slopes. As in any regression analysis, the coding and scaling of the between-person variable will lead to different estimates of effects. This type of SEM is depicted in Figure 23.4a and typically called by terms that are based on the history of its development with a substantive discipline: This model is termed an LCM in psychology (Meredith & Tisak, 1990), multilevel modeling in education (e.g., Bryk & Raudenbush, 1992), and a mixed-effects or random coefficients model in statistics (e.g., Longford, 1993). To be fair, there are advantages to considering these analyses within each discipline, but it is most important to note that variations of this analysis can be done using many computer programs, and the results are identical (see Ferrer, Hamagami & McArdle, 2004). In our prior WISC analyses, we have found that both the intercept and the slope were significantly higher for children whose mothers' educations were initially higher (see McArdle, 1989). Two of the limits of this popular approach quickly become evident: (1) We need to select an overall basis function for the entire group, and (2) the variance and covariance of the residuals of the latent variables must be the same over group.

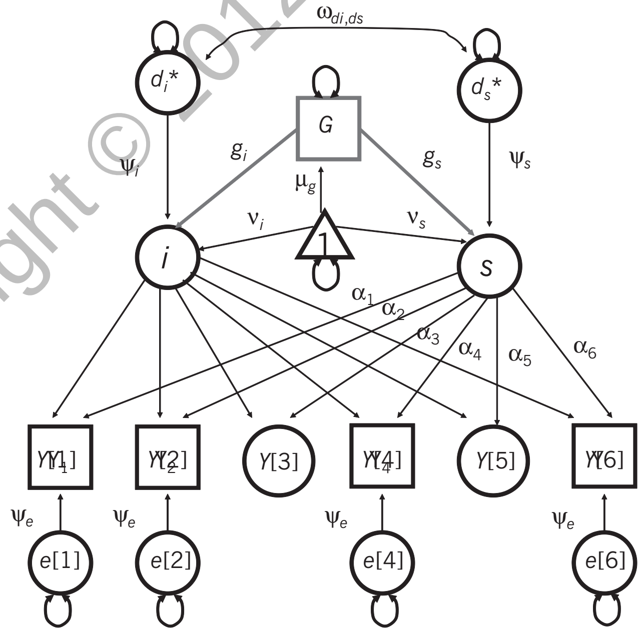

FIGURE 23.4a. A "multilevel" or "mixed-effects" model or LCM with an external G variable affecting both the latent levels and latent slopes.

A second analysis that does not require these constraints is to use the group variable to split the observed data into subsets for separate analyses. This is illustrated in the diagram of Figure 23.4b. This approach explicitly recognizes that separate models are allowed, so not all persons need to have the same within-person change model. In an SEM framework, checking whether or not the model and the model parameters are identical across independent groups is based on multiple group factor analysis (MGFA; Horn & McArdle, 1992; Sörbom, 1979).

Practical problems occur in any longitudinal study, including initial self-selection and dropout (or attrition). In prior work, a multiple-group SEM approach has been used (Jöreskog, 1971; McArdle & Hamagami, 1992) to deal with problems of "incomplete data" (see Cudeck, 2000; McArdle, 1994). In this situation, we can use the MGFA approach, defining the grouping variable based on patterns of available data to deal with incomplete pieces (with the incomplete data simply drawn as circles in the path diagram). We then assume that all parameters are invariant over all groups so that we mimic the assumptions of data missing at random (MAR; Little & Rubin, 1987; McArdle, 1994) and obtain tests of fit to represent MAR as a testable hypothesis.

In recent work much more complex models have been fitted using a raw data approach that does not require the data blocked into simple patterns (see Hamagami & McArdle, 2000; McArdle, Hamagami, Meredith, & Bradway, 2000; McArdle & Woodcock, 1997). Many of these can be seen from the perspective of multiple-group models with different numbers of data points on each person (McArdle & Hamagami, 1992). In this case, a model of "longitudinal convergence" (after Bell, 1954; McArdle & Bell, 2000) is a reasonable goal of many studies, but it is not always a hypothesis that can be tested with incomplete patterns (McArdle & Anderson, 1990). Instead, the MAR assumption is viewed as a convenient starting point, allowing us to use all the available information in one analysis, but it

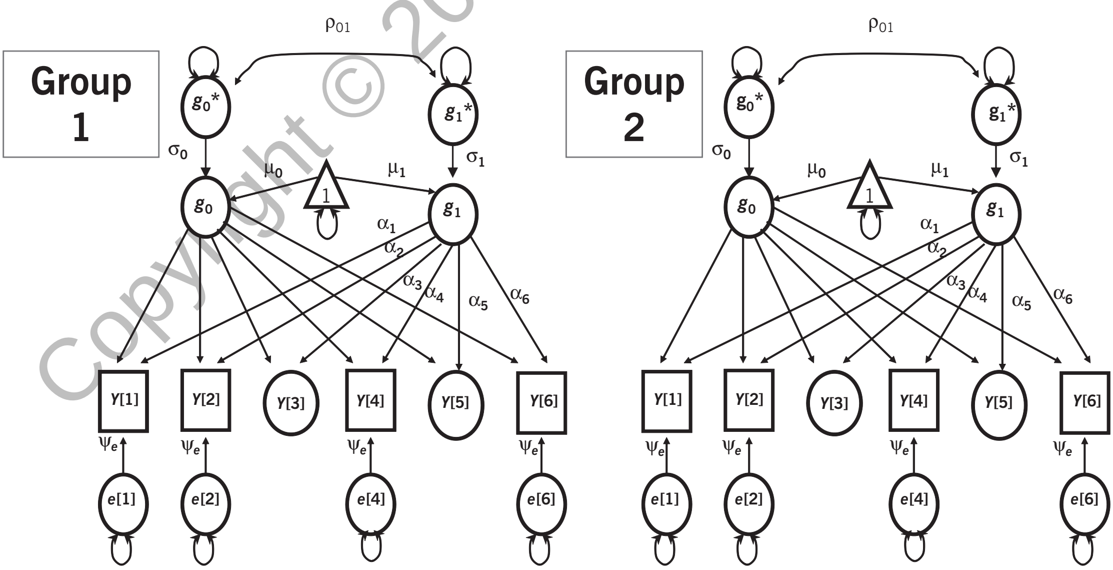

FIGURE 23.4b. Multiple-growth models for independent groups. The key question now is "Are any parameters not meaningfully the same (invariant) over groups?"

could be incorrect for a number of reasons (e.g., Cnaan, Laird, & Slasor, 1997; McArdle, 1994), but these key assumptions should be examined. Thus, instead of simply stating that MAR is a reasonable assumption, we should investigate methods for evaluating the failure of these assumptions. We can examine a model on the data that are already available, or we can gather extra information about the selection process (see Hedeker & Gibbons, 1997).

The benefits of considering incomplete group data are profound. For example, one of the corollary between-group issues is the potential for combining cross-sectional and longitudinal data. In one early resolution of these problems, Bell (1953) created a data collection he termed *accelerated* longitudinal data. To examine such ideas, Bell (1954) showed that by selecting data from a more time-consuming longitudinal study (i.e., by deletion of some of the existing data) he would get the same results. To ensure uniformity of the outcomes, Bell suggested a test of "convergence" of the cross-sectional and longitudinal data, created by joining together the shorter spans of data from people who were most similar at the endpoints. In contemporary work, this is now fitted with "invariance of the basis coefficients" (McArdle & Anderson, 1990). Bell (1954) demonstrated that estimates and inferences were basically the same, even though the data collection was accelerated, and we now do the same analysis using statistics for incomplete data (Hsiao, 2003; McArdle & Bell, 2000; Miyazaki & Raudenbush, 2000).

A fourth way of dealing with between-group differences is to create latent classes of different people based on the pattern of available data. It is always worth considering whether we have unmeasured subgroupings of people in our data set. It is now popular to explore these ideas with unobserved but heterogeneous subgroups, as in the SEM work on latent mixture models (LMM; Bauer & Curran, 2003; McLachan & Peel, 2000; Muthén, 2004, 2006). The defining feature of this model is that an SEM may be appropriate for each subgroup of persons, but there may be completely different SEMs for different unobserved subgroups; what we end up within the total group is a "mixture" of people. To better understand the class assignment of individuals, we can also write a prediction model (typically based on logistic regression) in which the probability of latent class membership is describable by other measured variables. The use of multiple parameters and a combined model is relatively new to this area, but it is potentially a reasonable and convenient way to deal with the problem of heterogeneity in a single-group data collection design. In the LCM context, this problem is often termed a "latent growth mixture model" (LGMM); this is essentially an effort to estimate the influences of LVs that are categorical, and this has an interesting history (see Bartholomew, 1987; Muthén & Shedden, 1999). This organization of information is the same as in Figure 23.4b, but, most important, the group membership is not known until the analysis is carried out. We have used LGMM with the WISC data and found this somewhat novel approach to be informative.

Issue 3: Studying Interrelationships in Individual Changes with SEM

The next question we face is what to do when we have multiple outcome variables. Figure 23.5a is a simple example of data from the WISC study. Here we plot both the longitudinal trajectory of the Verbal scales and the same for the Non-Verbal scales. The questions we ask begin with a surface look at the similarity of models across variables, but the deeper questions emerge about the time-dependent relationships of each variable to one another.

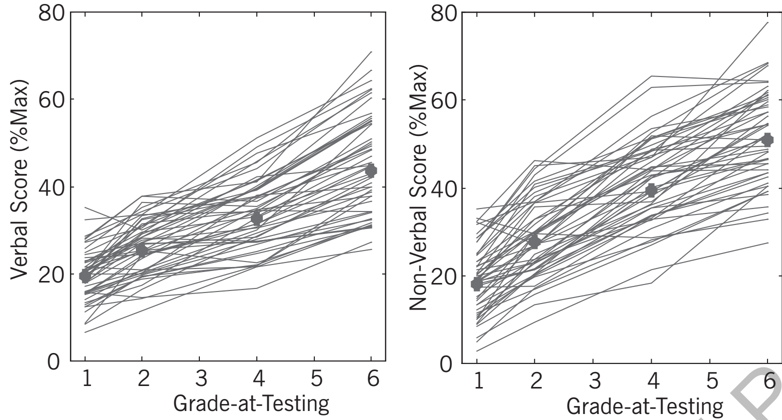

FIGURE 23.5a. Longitudinal WISC data on two variables ($N = 204$, plotted $n = 50$).

One natural model for multivariate longitudinal data is to examine the LCM of a common factor score. Basically, each MV is thought to be an indicator of one or more LVs, and it is this LV that carries the information over time. The key requirement here is that the factor pattern for the LV relationships is invariant over time (see McArdle, 1988, 2007). If invariance can be established, the LCM can then be applied to the LV[*t*] for examination of the common intercepts and slopes. In this model, the unique variance at the second level does not have any measurement error, so we have a clear indication of within-time variability that is not correlated over time (termed *state variance* by Cattell, 1959, as reported by McArdle & Woodcock, 1997). This model was termed a *curve of factor scores* (CUFFS; McArdle, 1988), and it still seems useful to some researchers (see Ferrer, Balluerka, & Widaman, 2008).

A different model for these same data is presented in Figure 23.5b. This model starts by fitting separate latent curves to each variable, and here we could establish any similarity of the curves—that is, we could first test the equality of the group slope parameters ($\alpha_1[t] = \alpha_2[t]$, etc.). Equality or invariance of this process is not required but can be studied. Most important, we can join these models at the individual level by correlating the latent intercepts and the slopes (see McArdle, 1989). That is, rather than correlating the variables at the same times, we correlated the components of the curves over time. The correlation of two or more latent slopes offers a broad single indication of whether one variable is changing in the same people as another variable. Testing the similarity of individual changes using the slope scores has proven to be a bit tricky, because the variances and covariances of the slopes need to be considered, and one is left wondering what to do about the intercept term. Basically, these are not separable concepts in most observational studies, because the intercepts and slopes are thought to occur at the same times. This SEM could be useful in the beginnings of an investigation into "correlated changes." If the correlation of components of change is our goal, we do not even need to measure the variables at exactly the same occasions or with exactly the same frequency.

One obvious extension of this logic is to examine common factors of the multiple intercepts and multiple slopes. One persistent problem with this model is the rather artificial separation of intercepts from slopes, because they are conceptually part of the same

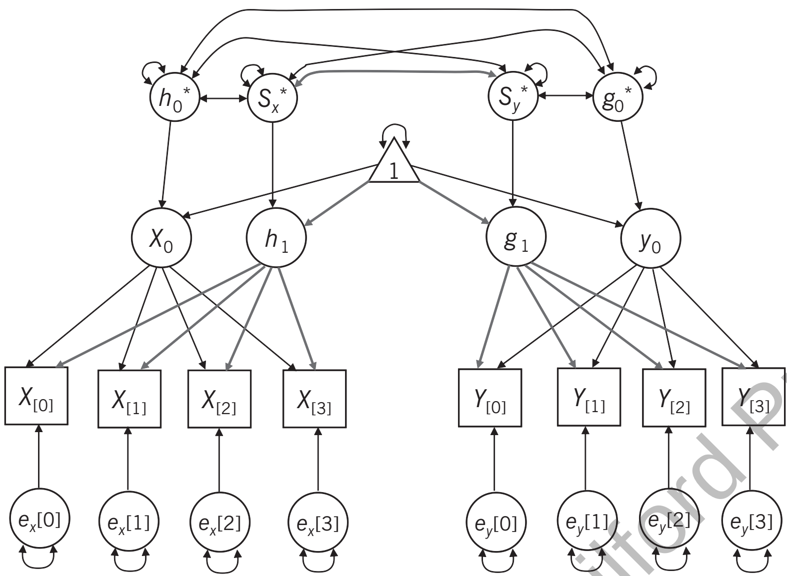

FIGURE 23.5b. A latent growth model for multiple variables. Based on McArdle (1988, 1989).

process. Nevertheless, such a model was initially termed a *factor of curves scores* (FOCUS; McArdle, 1988), and it is still considered useful by some researchers (see Duncan, Duncan, & Stryker, 2006).

No doubt there are many other models for multivariate longitudinal data that can be created to deal with this type of problem. However, most of these do not deal with specific timings, which was a direct focus of our autoregressive starting point. The bivariate and multivariate illustrations presented here dealt with the critical problem of interrelationships among different growth variables. In previous work, McArdle (1988, 1989, 1991) outlined some of these issues and fit a model in which the levels and slopes of one series was correlated with, or regressed upon, the levels and slopes of a different series. Similar multiple-variable latent variable modeling analyses have been used and reported by others (e.g., Raykov, 1997; Walker, Acock, Bowman, & Li, 1996). These kinds of analyses formalize an idea that the correlation among the slopes of two different variables may reflect the "common changes" (Griffiths & Sandland, 1984; McArdle & Nesselroade, 1994).

Issue 4: Studying Determinants (Causes) of Individual Changes with SEM

There are many alternative longitudinal models that can be fitted to these kinds of longitudinal data in Figure 23.5a (page 397), and some of these take into account the ordering of the responses. In these SEMs it is also important to recognize that models of *individual-differences relationships among dynamic variables do not necessarily reflect dynamic processes for the group*. The latter interpretation requires a focus on time-based parameters in the models and not on the time-independent correlation of the time-based scores (e.g., Gollob & Reichardt, 1987).

Following the previous work on common factor analysis and cross-lagged regression (Jöreskog & Sorbom, 1979) we can simply assume that, for the MV[*t*], the observed raw score measured can be decomposed into the sum of a common factor score and a unique score. This algebraic form is depicted in the path diagram of Figure 23.6a by having the latent variables and the error scores add to each other (i.e., with unit weights) to form the observed variables. Additional restrictive assumptions about random errors imply that the true scores include any common factors, as well as any specific factors.

Next we need to represent the changes within a person over time. We define a latent change score (LCS) between any two latent scores as the part of a later latent score at one time that is not part of it at the earlier time. We can simplify a lot of the required algebra by always assuming that the interval of time between LVs is the same. The simplifying assumption of equal latent intervals is not a testable feature of the model. In practice, it is not always possible and not always desirable to keep the MV intervals of time equal (see McArdle & Woodcock, 1997, 2000). We deal with this practical problem by making a clear definition of the patterns of complete and incomplete data (see Horn & McArdle, 2007; Jöreskog & Sorbom, 1979; McArdle, 1997; McArdle & Aber, 1990; McArdle & Bell, 2000; McArdle & Hamagami, 1991). In the path diagram of Figure 23.6a the LCS is drawn as a latent variable at each time, and the interpretation of this score as a change within a person is assured by the placement of the latent score at the first occasion and the LCSs (see McArdle & Nesselroade, 1994).

We can now create an SEM that allows for dynamic inferences and reinforces the key idea—we now have individual changes represented as LVs, and these are the key outcomes. What the researcher must now do is pose a model for the latent changes (see Ferrer & McArdle, 2010), and this can be difficult for a number of reasons. First, we are not used to

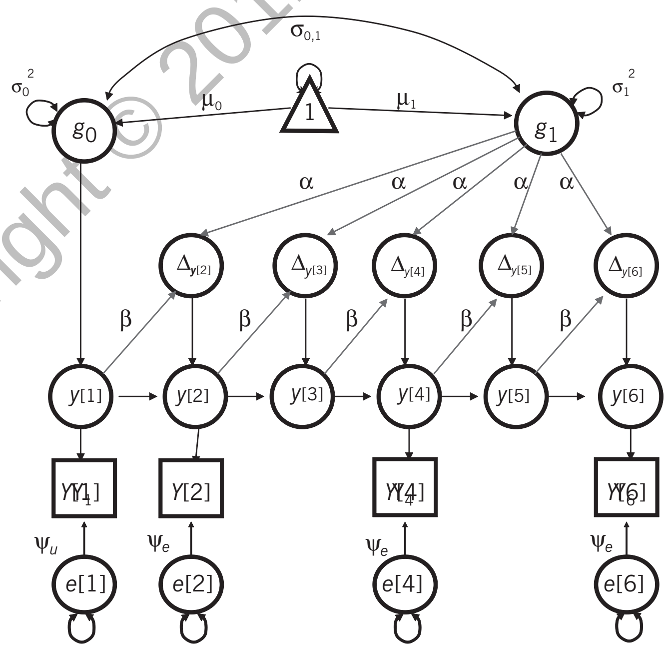

FIGURE 23.6a. The latent change score (LCS) model based on both LCM and AR(1).

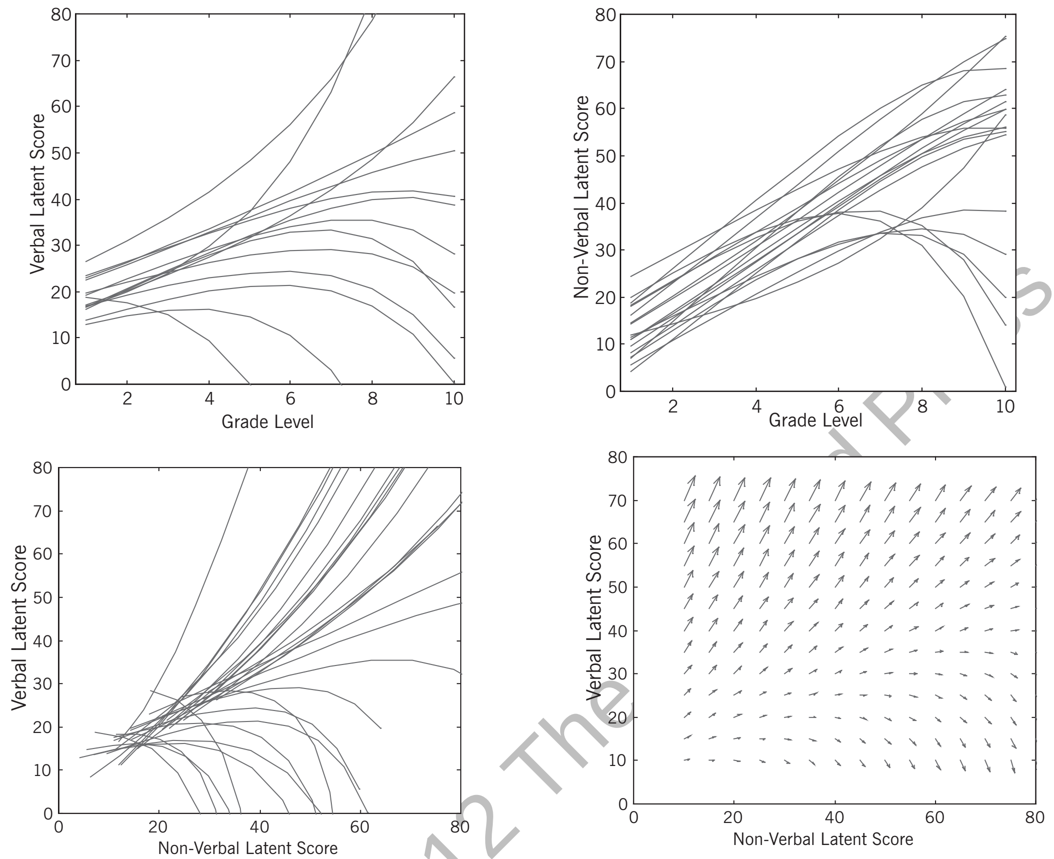

FIGURE 23.6b. Expected bivariate latent trajectories.

doing this kind of work on an a priori basis, but this is needed to create an expected set of trajectories. Second, we are typically willing to explore and try a lot of different models to see what looks like the best fit. In the SEM approach used here, we are not restrictive, and many models of change are possible.

These choices of dynamic models may be made on substantive or empirical considerations. In our initial examples, the model of latent change we used (Figure 23.6a) was based on a simple additive combination of the linear curve model and the autoregressive model (Figure 23.2, pages 388–389) and the linear latent curve model (Figure 23.3, page 391). This was written so time-based changes in the latent change score were allowed (1) to be constantly related to some alternative but fixed slope score ($\alpha[t]\, f_{1n}$), (2) to be proportional to the previous state ($\alpha[t]\ .\ y[t-1]_n$), and (3) to include an additive residual or disturbance term ($z[t]$). This dynamic SEM can be completed by adding several parameters for the multiple sets of latent variables. In all cases, we use fixed unit values to define the LCS, and we ensured the LCS were outcome variables. That is, the earlier LV in time is included as a predictor of the LCS, as is the common latent slope. All dynamic coefficients are repeatedly used for all occasions, but this is a hypothesis that should be tested. Due to our use of two additive change components, we termed this a *dual change score* model (DCS; McArdle, 2001; McArdle & Hamagami, 2001). This DCS model has some interesting properties, including its ability to bring together both the LCM and the AR–SEM as proper statistical subsets. We can assume that the latent trajectories are smooth deterministic curves or have uniform random residual properties by eliminating the variance of the residual ($z[t]$).

These kinds of univariate models were initially represented for the Verbal (V) and Non-Verbal (NV) scales of the WISC and were fitted using LISREL software. The results are presented graphically elsewhere (see McArdle, 2001). In each case (of V and NV) the models required at least as many parameters as were available in the DCS framework. That is, neither the autoregressive SEM nor the LCM fit these data very well, but the DCS was a great improvement. The resulting DCS equation for these trajectories for the overall group curve from the model parameters shows that the expected group mean will always increase both additively (by +2) and proportionally (by +0.09) per year. The simple substitution of specific time points ($t = 0$–10) into this equation results in the nearly straight average group curve. We note that trajectory estimates are made for grade 3 and grade 5, even though there were no group observations at this point, and also for $t > 6$ as a "forecast" of scores in future years. To better understand the meaning of the individual-differences parameters, individual latent curves were generated from the repeated application of the difference equation, and this yielded individual curves that follow the particular solution. The resulting individual latent curves (for $n = 50$) are presented elsewhere and show significantly more nonlinear curvature than the means (see McArdle, 2001). This collection of curves also shows increasing variance of a "fan spread" (Shadish, Cook, & Campbell, 2002).

These graphic illustrations from the WISC show only a sampling of the family of curves that can be generated from the few DCS model parameters. The results for the V model do not seem the same as the results for the NV; most obviously, one seems to grow up (V) and the other (NV) flattens out during this period of time. Tests for these kinds of bivariate models are well known in the SEM literature. A more challenging set of models is based on "time-varying covariates," or "systems models," and these are often not fully specified in SEM terms. However, let us now consider a bivariate model of change in which we write the prior difference equation for more than one dynamic variable.

If we want to identify the interplay of two variables that are changing over time, we can follow the previous SEM logic and write the LCS model in which the change in one latent variable is a function of both itself and another variable. The most important features of this bivariate change SEM include (1) the separation of individual scores from group parameters, (2) the assumption of a constant time interval, and (3) a separation of the true scores from the measurement error scores. Residual disturbances for each variable are included as well, but these will not be used. A path diagram of this bivariate dynamic model is presented in Figure 23.7a. As part of this SEM we include a representation of the change over time in the second LCS, including a constant change (α_x), a proportional effect (β_x) of the latent variable upon itself, and a "coupling effect" (γ_x) of the first variable. In this dynamic SEM it is clear that the predictors of one latent variable are embedded in the outcomes of the other variable and vice versa. In this bivariate form, model parameters may be simplified to deal with dynamic hypotheses about "coupling" or "lead-lag" relationships. Unlike all prior SEMs, these dynamic hypotheses may be considered in the presence of individual differences in constant growth and change. These alternatives can be examined in combination with restrictions on the univariate (μ and σ) parameters of change within each variable, and the results can be plotted as a statistical vector field (SVF; Boker & McArdle, 2005; McArdle & Hamagami, 2001).

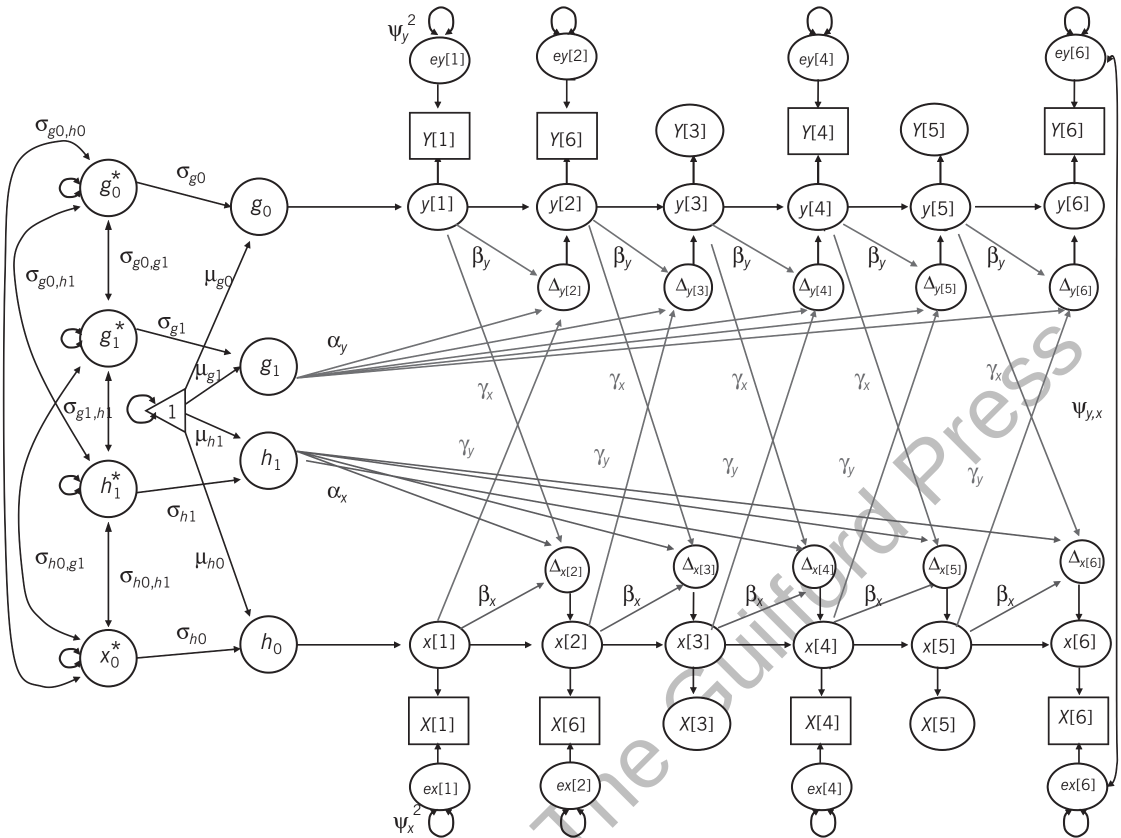

FIGURE 23.7a. A bivariate dual change score model.

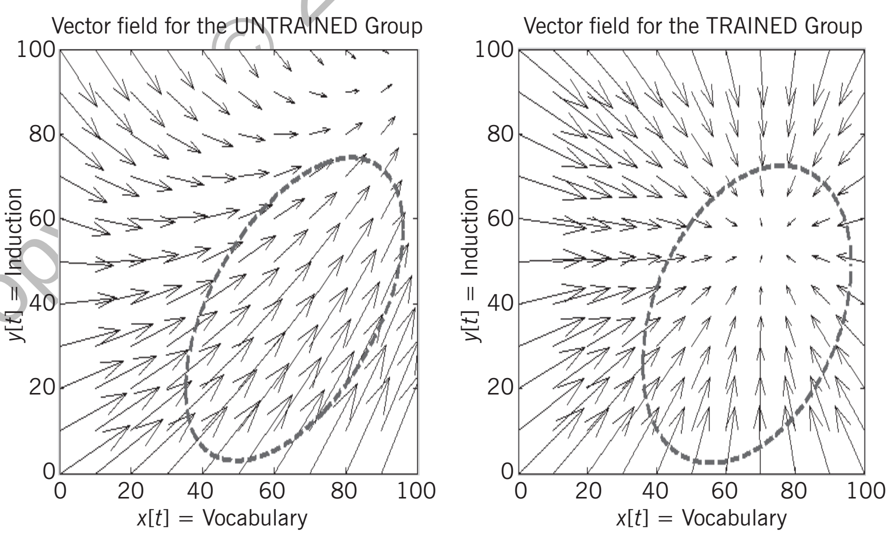

FIGURE 23.7b. Vector field plots for untrained and trained groups from the Berlin Aging Study of Training. Ellipse is 95% confidence around starting points.

The numerical results of fitting two bivariate DCS models to the four-occasion WISC V and NV scores (of Figure 23.5a, page 397) have been found using standard SEM programs, and some results are presented graphically in Figure 23.6b (page 400). In contrast to the univariate LCD results for V scores, here we obtained a large auto-proportion ($\alpha_y = +.5$), a small but negative mean slope ($\beta_{sy} = -1$), and a nontrivial coupling parameter from latent NV[t – 1] to LCS[V] ($\gamma_y = -.2$). In contrast, the model for NV scores includes a smaller autoproportion ($\gamma_x = -.0$), a smaller positive mean slope ($\mu_{sx} = 5$), and only a trivial coupling. These bivariate results were used to calculate the expected values for a longer time sequence (t = 0–10). The plot of the V latent scores now shows what appears to be a rapid fanning out at both the upper and lower portions of the initial scores, whereas the NV scores appear to have more steady increases. The bivariate equations lead to latent curves with downward trends for both variables, with a larger impact on the V than the NV curves.

The final two panels of Figure 23.6b represent a different kind of time-sequence plot because here, for each individual, we draw the V latent curve scores paired against the NV latent curve scores for the same occasions. This "state-space" pairing of latent growth curves seems to show that most individuals follow rapid increases in both variables with a few individuals dropping off toward lower V scores but higher NV scores. The initial levels and slopes are all positively correlated (e.g., $\rho_{sy,sx} = .8$), and these correlations are informative because they describe the location of the individual dynamic curves in the time-sequence plots. The final panel of Figure 23.6b is the SVF version of the same bivariate curves. Each arrow here shows the general direction of all curves within that specific region (equal-sized cells) of these curves. That is, for any pair of latent scores at a specific time (i.e., $x[t]$, $y[t]$), the small arrow points to where the pair of latent scores is expected to be at the next time ($x[t + 1]$,$y[t + 1]$). This final figure seems to emphasize the increases in the upper left-hand quadrant of the bivariate field.

A model with no effect of V on the NV changes was evaluated with a restriction that V[t – 1] did not affect LCS[NV], and this alternative fit the WISC data fairly well. In addition, a model with no effect of NV on the V changes was fit where the NV[t – 1] did not affect LCS[V], but this did not fit the data. The collection of bivariate latent change score results leads to a few substantive conclusions about this dynamic system: (1) The variables are growing, so the standard cross-lagged assumptions (i.e., no growth) are not a viable alternative for these WISC data; (2) there is no effect of the Verbal latent scores on the NV latent changes; and (3) the NV latent scores significantly and negatively affect the V latent changes. Thus *the NV scores are leading indicators of the growth in V scores, but not vice versa*. As it turns out, this result is consistent with aspects of some prior developmental theory in this area (see McArdle, Hamagami, Meredith, & Bradway, 2000). There are not many other SEMs that are used to establish time-ordered precedence (see Ferrer & McArdle, 2010; McArdle, 2009).

Issue 5: Studying Individual Differences in the Determinants of Individual Changes

Our final goal is to combine the concepts of the previous models but add separate group issues. One way to see this is as a multiple-group (MG) SEM, and this might start with the bivariate change model in Figure 23.7a. Without further detail, we should be able to see this path diagram for multiple groups (as in McArdle & Prindle, 2008). In any case, all forms of between-group analysis are of obvious importance in the study of dynamics: (1) MVs that are measured at one time (static) can be added into the dynamic model, and this could change the coupling parameters; (2) the MG–SEM approach allows all of these options, plus it allows the dynamic influences to be completely different across different manifest groups; (3) incomplete data modeling by MG–SEM or other approaches is fundamental in dynamic analysis (see Hamagami & McArdle, 2000); and (4) the determination of unobserved latent classes of persons who are following the same dynamic process is essential (see McArdle & Prindle, 2008). Variations on these MG–SEMs can be used to examine different combinations of variables or latent mixtures (Muthén & Muthén, 2005; Nagin, 1999).

The methods and results of research on group differences in dynamics have not been applied to these WISC data. But these kinds of SEMs are detailed in McArdle (2006) with data from the Berlin Age Training Study (with the permission of the authors, Baltes, Dittmann-Kohli, & Kliegl, 1986). The general purpose of this data collection was to evaluate theory of "cognitive plasticity," which suggested that older adults have untapped cognitive resources that can be stimulated by cognitive training. The study participants were randomly assigned to one of two groups: (1) one in which they were trained on one of two cognitive tasks ($n = 161$) or (2) a no-contact (not trained) control group ($n = 87$). The cognitive training program consisted of 10 consecutive sessions, and each dealt with training specifically focused on the improvement of aspects of "fluid intelligence" or "reasoning in novel situations" using training methods developed in earlier research. A cognitive measurement design included four rounds of repeated cognitive testing: ($t = 0$), a pretest measurement followed by 1 month of training or of no contact; ($t = 1$), a 1-week posttest measurement; ($t = 2$), a 1-month posttest measurement; ($t = 3$), and a 6-month posttest ($t = 4$) measurement. These time points were chosen because they were thought to represent critical junctures in the learning of these cognitive tasks. At each measurement session, a battery of cognitive tests was repeatedly administered, and two of these tests were used here: (1) Induction (IND), a "near-transfer" test expected to be directly affected by the specific training; and (2) Vocabulary (VOC), a "far-transfer" test not expected to be affected by this training.

Standard repeated-measures ANOVA models were used by Baltes and colleagues (1986), but these analyses were not intended to capture the possible within-persons changes, or individual transfer effects. Raykov (1997) reanalyzed one subset of these data using a model for multiple indicators of an Induction factor. The SEMs were based on factor analysis in multiple groups with correlated growth curves (McArdle, 1989). Raykov (p. 310) concluded, "the results indicate (a) group equivalence in the pattern of temporal development of ability variances, and (b) training effects in the experimental group that are stronger than the practice/experience effects in the control group, and both types of effects are maintained over the 6-month testing period." McArdle (2007) used the same data but asked the specific questions about group differences due to training as a dynamic system. The MG–SEM approach allows for differences in the latent means, slope variances, residual variance, and dynamic coupling of the respective trajectories. But the groups could have a different set of coupling parameters, implying a more fundamental impact of the experiment—*alterations in the dynamic system due to the treatment*.

These new results suggested that the Vocabulary scores exhibit proportionally more unique variance (and hence are less reliable, although the correlation among the latent levels is positive [+.5]). In terms of the group differences, one clear training impact is that the constant slopes are much higher in the trained group for both variables. This is interesting as the second variable (Vocabulary) is a "far-transfer" variable, and no real training differences were expected. This interpretation of mean changes for different variables is often characterized as a "transfer of training" (e.g., Baltes et al., 1986). The correlations among the latent slopes are high (+.8) in the control group and lower in the experimental group (+.5), but this is not what we expected as an impact when training is transferred. Further group differences include parameters with different signs, so their total impacts are hard to interpret directly or individually. In the control group, all parameters were accurately different from zero, so this was a *coupled system* of latent scores. In contrast, the results for the treated group suggested higher autoproportions and no accurate coupling parameters, so this was as an *uncoupled system* of latent scores.

The net result of the experimental training manipulation was that the treatment both raised and uncoupled the scores on the two variables. This result suggests no strong evidence of a transfer of training across these variables, This result could be due to the fact that training helped focus the persons on all tasks and especially focused on the Induction-like tasks (near transfer). At the same time the pattern of individual differences apparent here suggests that the specific training seems to have set the Induction variable farther away from the Vocabulary variable than it was before. The expected values from these parameters are those displayed as *vector field plots* for each group in Figure 23.7b (page 402). Both figures show an interesting dynamic property—*the change expectations of a dynamic model depend on the starting point*. Thus, the training seems to have isolated the abilities that were trained and possibly moved these scores toward other, unmeasured variables.

It may be clear that these kinds of dynamic differences due to training are not easy to see in the typical comparison of changes in the means, as in repeated-measures ANOVA. An updated version of this kind of MG–SEM dynamic analysis has also been carried out for a different set of randomized experimental data in McArdle and Prindle (2008) and for observational data in Grimm and McArdle (2007).

Discussion: Studying Longitudinal Data with Contemporary SEM

This analysis of longitudinal trajectories can be a great deal of fun for the eager analyst. One has an overwhelming sense that something useful, something special, could emerge from analyzing longitudinal data. After all, almost any longitudinal data collection takes a lot more time and effort than the more typical cross-sectional data collection. So, whether or not it is deserved, longitudinal analyses often take on a special importance. But, for very similar reasons, these analyses can be very difficult. Longitudinal analyses can be difficult to organize. It can be difficult to do something very simple and useful (such as plotting the data) and difficult to know when to stop with advanced analyses. To deal with these issues, I have placed SEM in the context of five goals suggested by Baltes and Nesselroade (1979). I suggest you consider what you or others have done within the context of the five steps outlined earlier; it is not necessary to carry them out in order, but it could be easier. Each step can build on the results from the previous steps.

There are many motivations to use contemporary SEM for longitudinal analyses, and SEM has both benefits and limitations. Among many benefits, we can see a required clarity of process definitions, easy programming of complex ideas, clear tests of parameter equivalence, and generally simple tests of otherwise complex ideas. A few of these SEM techniques even lead to novel ideas for data analysis, especially dealing with ordinal measurements. Perhaps SEM path diagrams provide a clearer way to think about longitudinal analyses. Unfortunately, among many SEM limitations, we need large and representative samples of persons for reasonable estimation, and we also need normal distribution of residuals (or uniquenesses) to create appropriate statistical tests. But, most important, we need to state what we want to examine about dynamic influences in advance of the data analysis. In this way any SEM requires the kind of clear thinking we have probably not yet achieved in behavioral sciences. In common practice, we get stuck using the SEM models in ways that are far simpler than is possible. That is, we seem to focus on issues 1, 2 and 3 and rarely reach issues 4 and 5.

Once again, the choice of a dynamic model is intimately tied to the substantive theory and the available data. The choice between using an autoregressive approach, a linear latent basis, or a polynomial or other nonlinear model can be substantively important. The initial idea from Wishart (1938) was that the basic shape of each individual curve could be captured with a small number of fixed parameters and random variance components. In some cases a fixed-basis polynomial model is more parsimonious than a free-basis model (see McArdle & Bell, 2000). But many recent textbooks overlook the latent-basis model (e.g., Duncan et al., 2006; Singer & Willett, 2003), and this option remains a most useful empirical choice.

One way to combine recent ideas about dynamics and retain the statistical and practical benefits of SEM is to use a model based on the LCS (Hamagami & McArdle, 2000; McArdle, 2001; McArdle & Hamagami, 2001). Indeed, the introduction of individual differences in change analyses has led to a great deal of statistical controversy in model fitting (e.g., Nesselroade & Cable, 1974; Sullivan, Rosenbloom, Lim, & Pfefferman, 2000). These kinds of "difference" or "gain" score calculations are relatively simple and theoretically meaningful, but the potential confounds due to the accumulation of random errors has been a key concern in previous studies using observed change scores or rate of change scores (e.g., Burr & Nesselroade, 1990; Rogosa & Willett, 1985; Willett, 1990). This recent SEM approach allows us to write expectations in terms of the first- or second- or higher order latent differences, and we can have the SEM program automatically generate the required nonlinear expectations of the scores.

Practical problems in the fitting of any statistical model with longitudinal data begin with scaling and metrics. These ideas can be examined at the item level by forming a scoring system for any construct; we used concepts from item response theory (IRT) (see McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009; McArdle & Hamagami, 2004). Additional scaling issues deal with the exact timing of observations (e.g., Boker & McArdle, 2005) and, as usual, transformations can change the statistical patterns. Optimal measurement collections are a major challenge for new empirical research and worthy of much further discussion (McArdle, 2010).

The SEM analyses presented here make specific assumptions about latent changes in individuals and groups (as in McArdle & Prindle, 2008; Meredith & Tiask, 1990; Muthén & Curran, 1997). But the SEMs presented here were developed to be a reasonable match to the *goals of longitudinal data analysis* proposed by Baltes and Nesselroade (1979; Table 1). The representation of change processes can be accomplished using the variety of methods we have been referring to as latent curve, mixed effect, multilevel, or multiple-group modeling. With this in mind, we need to be careful not to forget the important lessons of the past. At the same time, we do need to invent new ways to collect the optimal sets of data needed for powerful tests of hypotheses about development. These are among the most important challenges for contemporary and future longitudinal research, and my hope is that some of the SEM ideas presented here can be useful in this work.

Acknowledgments

This work was supported by NIA Grant AG-7137-21 to John J. McArdle. I thank my colleagues John R. Nesselroade (University of Virginia), Kevin Grimm (University of California at Davis), John Prindle (University of Southern California), and Todd Little (University of Kansas) for their important contributions to an earlier version of this work.

References

Andersen, T. W. (1971). *The statistical analysis of time series*. New York: Wiley.

Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserved capacity of the elderly in aging-sensitive tests of fluid intelligence: Replication and extension. *Psychology and Aging, 1*(2), 172–177.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. Gurnee, IL: Kenner Books.

Bauer, D. J., & Curran, P. J. (2003). Over-extracting latent trajectory classes: Much ado about nothing? *Psychological Methods, 8*, 384–393.

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24*, 145–222.

Bell, R. Q. (1954). An experimental test of the accelerated longitudinal approach. *Child Development, 25*, 281–286.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Boker, S., & McArdle, J. J. (2005). Vector field plots. In P. Armitage & P. Colton (Eds.), *Encyclopedia of Biostatistics* (2nd ed., Vol. 8, pp. 5700–5704). Chichester, UK: Wiley.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Advances in psychometrics: A festschrift to Roderick P. McDonald* (pp. 415–451). Mahwah, NJ: Erlbaum.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage.

Burr, J., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), *Statistical methods in longitudinal research* (Vol. 1). New York: Academic Press.

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*, 2349–2380.

Cudeck, R. (2000). An estimate of the covariance between two variables which are not jointly observed. *Psychometrika, 65*, 539–546.

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

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology, 4*, 22–36.

Ferrer, E., Hamagami, F., & McArdle, J. J. (2004). Modeling latent growth curves with incomplete data using different types of structural equation modeling and multilevel software. *Structural Equation Modeling, 11*(3), 452–483.

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science, 19*(3), 149–154.

Ferrer-Caja, E., & McArdle, J. J. (2004). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 10*(4), 493–524.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80–92.

Griffiths, D., & Sandland, R. (1984). Fitting generalized allometric models to multivariate growth data. *Biometrics, 40*, 139–150.

Grimm, K. J., & McArdle, J. J. (2007). A dynamic structural analysis of the potential impacts of major context shifts on lifespan cognitive development. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 363–386). Mahwah, NJ: Erlbaum.

Hamagami, F., & McArdle, J. J. (2000). Advanced studies of individual differences linear dynamic models for longitudinal data analysis. In G. Marcoulides & R. Schumacker (Eds.), *New developments and techniques in structural equations modeling* (pp. 203–246). Mahwah, NJ: Erlbaum.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*, 64–78.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley-Interscience.

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. MacCallum (Eds.), *Factor analysis at 100 years* (pp. 205–247). Mahwah, NJ: Erlbaum.

Hsiao, C. (2003). *Analysis of panel data*. New York: Cambridge University Press.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Jöreskog, K. G., & Sorbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Little, R. T. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Longford, N. T. (1993). *Random coefficient models*. London: Clarendon Press.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*(1), 163–200.

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

McArdle, J. J. (1989). Structural modeling experiments using multiple growth functions. In P. Ackerman, R. Kanfer, & R. Cudeck (Eds.), *Learning and individual differences: Abilities, motivation, and methodology* (pp. 71–117). Hillsdale, NJ: Erlbaum.

McArdle, J. J. (1991). Structural models of developmental theory in psychology. *Annals of Theoretical Psychology, 7*, 139–160.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*(4), 409–454.

McArdle, J. J. (1997). Modeling longitudinal data by latent growth curve methods. In G. Marcoulides (Ed.), *Modern methods for business research* (pp. 359–406). Mahwah, NJ: Erlbaum.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 342–380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J. (2006). Dynamic structural equation modeling in longitudinal experimental studies. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioural and related sciences* (pp. 159–187). Mahwah, NJ: Erlbaum.

McArdle, J. J. (2007). Five steps in the structural factor analysis of longitudinal data. In R. Cudeck & R. MacCallum (Eds.), *Factor analysis at 100 years* (pp. 99–130). Mahwah, NJ: Erlbaum.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577–605.

McArdle, J. J. (2010). Contemporary challenges of longitudinal measurement using HRS data. In G.

Walford, E. Tucker, & M. Viswanathan (Eds.), *Sage handbook of measurement* (pp. 509–536). London: Sage.

McArdle, J. J., & Aber, M. S. (1990). Patterns of change within latent variable structural equation modeling. In A. von Eye (Ed.), *New statistical methods in developmental research* (pp. 151–224). New York: Academic Press.

McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 21–43). New York: Plenum Press.

McArdle, J. J., & Bell, R. Q. (2000). An introduction to latent growth curve models for developmental data analysis. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and scientific examples* (pp. 69–107). Mahwah, NJ: Erlbaum.

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

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal multilevel structural analyses of the growth and decline of multiple intellectual abilities over the life-span. *Developmental Psychology, 38*(1), 115–142.

McArdle, J. J., Grimm, K., Hamagami, F., Bowles, R., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods, 14*(2), 126–149.

McArdle, J. J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 276–304). Washington, DC: American Psychiatric Association.

McArdle, J. J, & Hamagami, F. (2001). Linear dynamic analyses of incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–176). Washington, DC: American Psychological Association.

McArdle, J. J, & Hamagami, F. (2004). Methods for dynamic change hypotheses. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments in structural equation models* (pp. 295–336). London: Kluwer Academic.

McArdle, J. J., Hamagami, F., Meredith, W., & Bradway, K. P. (2000). Modeling the dynamic hypotheses of Gf–Gc theory using longitudinal life-span data. *Learning and Individual Differences, 12*(1), 53–79.

McArdle, J. J., & Nesselroade, J. R. (1994). Using multivariate data to structure developmental change. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological innovations* (pp. 223–267). Hillsdale, NJ: Erlbaum.

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analyses in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology: Vol. 2. Research methods in psychology* (pp. 447–480). New York: Pergamon Press.

McArdle, J. J., & Prindle, J. J. (2008). A latent change score analysis of a randomized clinical trial in reasoning training. *Psychology and Aging, 23*(4), 702–719.

McArdle, J. J., & Woodcock, J. R. (1997). Expanding test-rest designs to include developmental time-lag components. *Psychological Methods, 2*(4), 403–435.

McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development, 44*, 401–415.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

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

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods, 5*(1), 24–63.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B. (2006). The potential of growth mixture modeling. *Infant and Child Development, 15*, 623–625.

Muthén, B., & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*, 371–402.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463–469.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus, the comprehensive modeling program for applied researchers user's guide*. Los Angeles: Authors.

Nagin, D. (1999). Analyzing developmental trajectories: Semi-parametric group-based approach. *Psychological Methods, 4*, 139–177.

Nesselroade, J. R., & Cable, D. G. (1974). "Sometimes it's okay to factor difference scores": The separation of trait and state anxiety. *Multivariate Behavioral Research, 9*, 273–282.

Osborne, R. T., & Suddick, D. E. (1972). A longitudinal investigation of the intellectual differentiation hypothesis. *Journal of Genetic Psychology, 121*(1), 83.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501–525.

Raykov, T. (1997). Growth curve analysis of ability means and variances in measures of fluid intelligence of older adults. *Structural Equation Modeling, 4*, 283–319.

Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Mahwah, NJ: Erlbaum.

Rogosa, D., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*(2), 203–228.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

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

Sörbom, D. (1979). An alternative to the methodology for analysis of covariance. *Psychometrika, 43*, 381–396.

Sullivan, E. V., Rosenbloom, M. J., Lim, K. O., & Pfefferman, A. (2000). Longitudinal changes in cognition, gait, balance in abstinent and relapsed alcoholic men: Relationships to changes in brain structure. *Neuropsychology, 14*(2), 178–188.

Wainer, H., & Spence, I. (2004). Graphical presentation of longitudinal data. In B. S. Everitt & D. C. Howell (Eds.-in-Chief), *Encyclopedia of statistics in behavioral science* (pp. 1–10). New York: Wiley.

Walker, A., Acock, A. C., Bowman, S. B., & Li, F. (1996). Amount of care given and caregiving satisfaction: A latent growth curve analysis. *Journal of Gerontology: Psychological Science and Social Science, 51B*(3), 130–142.

Walls, T. A., & Schafer, J. L. (Eds.). (2006). *Models for intensive longitudinal data*. Oxford, UK: Oxford University Press.

Wechsler, D. (1970). *The Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.

Willett, J. B. (1990). Mearsuring change: The difference score and beyond. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of education evaluation* (pp. 632–637). Oxford, UK: Pergamon.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Wishart, J. (1938). Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika, 30*(1/2), 16–28.

Wohlwill, J. (1973). *The study of behavioral development*. New York: Academic Press.