Psychology and Aging
2003, Vol. 18, No. 4, 639–657

0882-7974/03/$12.00 DOI: 10.1037/0882-7974.18.4.639

Assessing Psychological Change in Adulthood: An Overview of Methodological Issues

Christopher Hertzog
Georgia Institute of Technology

John R. Nesselroade
University of Virginia

This article reviews the current status of methods available for the analysis of psychological change in adulthood and aging. Enormous progress has been made in designing statistical models that can capture key aspects of intraindividual change, as reflected in techniques such as latent growth curve models and multilevel (random-effects) models. However, the rapid evolution of statistical innovations may have obscured the critical importance of addressing rival explanations for statistical outcomes, such as cohort differences or practice effects that could influence estimates of age-related change. Choice of modeling technique and implementation of a specific modeling approach should be grounded in and reflect both the theoretical nature of the developmental phenomenon and the features of the sampling design that selected persons, variables, and contexts for empirical observation.

At the heart of the science of aging is the requirement that one have appropriate concepts and methods for evaluating how and why individuals change (or remain stable) as they grow older (Baltes, Staudinger, & Lindenberger, 1999; Nesselroade, 1991; Wohlwill, 1973, 1991). The last 40 years has witnessed considerable progress in our understanding of how to conceptualize and measure psychological change, as exemplified by a substantial number of reviews and books dedicated to the topic (e.g., Baltes & Nesselroade, 1979; Collins & Horn, 1991; Collins & Sayer, 2001; Little, Schnabel, & Baumert, 2000). Our review provides a perspective on the progress made to date and touches on the prospects for future advances. We selectively review these technical advances in terms of their relevance for research on psychological aging.

First, we briefly cover basic concepts that are important for understanding the measurement of change. Second, in the bulk of the article, we review (in conceptual rather than mathematical terms) aspects of key statistical methods for the measurement of change in adulthood. Emphasis in the recent literature has been placed on the problem of statistical methods for assessing intraindividual change (Collins & Horn, 1991; Collins & Sayer, 2001), resulting in more widespread application of techniques like multilevel models (MLMs) to longitudinal data. Moreover, new methods are emerging that, although not yet in widespread use, promise to radically alter concepts about the dynamics of change in adulthood (Boker, 2001; Nesselroade & Molenaar, 2003). Thus, the field is rapidly developing coherent and sensible approaches to estimating intraindividual change and individual differences in rates of change. Third, we consider the relevance of concepts of adult development and methodological issues about designing studies to measure development for our thinking about statistical models for adult development and aging. In our view, a cost of the emphasis on statistical advances is diverted attention from the problem of matching theoretical conceptions about adult development with design, measurement, and analysis decisions when modeling change (Alwin & Campbell, 2001). We believe developmental psychologists can benefit from revisiting issues about impact of research designs and assumptions on the validity of the substantive inferences about developmental change (Baltes, Reese, & Nesselroade, 1988; Hertzog & Dixon, 1996; Horn & Donaldson, 1976; Schaie, 1977).

Measurement of Change

It borders on a truism that without ways to conceptualize and measure change we cannot have a scientific study of aging. Even though the representation and measurement of change is a longstanding topic in behavioral science, there has been a noticeable lack of agreement among the experts on measuring, representing, and conceptualizing quantitative change (for divergent perspectives, compare Bereiter, 1963; Cronbach & Furby, 1970; Hummel-Rossi & Weinberg, 1975; Nesselroade, 1991; and Wohlwill, 1991). Hindsight reveals an interesting bifurcation in the history of change measurement in social and behavioral science. While the experts were arguing about whether change could be measured and just how it should be done, researchers proceeded with the measurement of change, using whatever methods were available, de-

Editor's Note. Leah L. Light served as the action editor for this article.—LLL

Christopher Hertzog, School of Psychology, Georgia Institute of Technology; John R. Nesselroade, Department of Psychology, University of Virginia.

John R. Nesselroade's work on this article was supported by National Institute on Aging Grant R01 AG18330 and by the Max Planck Institute for Human Development, Berlin, Germany.

For additional information, see Christopher Hertzog's World Wide Web page at http://psychology.gatech.edu/chertzog.

Correspondence concerning this article should be addressed to Christopher Hertzog, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, Georgia 30332-0170. E-mail: christopher.hertzog@psych.gatech.edu

spite, if not ignoring, the controversies. Perhaps erroneous inferences were generated in the process, but that seems no worse than sitting back and making no attempt to measure change—especially when changes are the basic phenomena of one's field.

Today, fundamental shifts are occurring in the way students of adult development and aging think about change and process (Nesselroade & McCollam, 2000). The assessment of behavioral and psychological change, rather than being a straightforward measurement problem, is best regarded as requiring a concerted effort to integrate relevant principles across the domains of measurement, research design, and statistical modeling (Nesselroade & Ghisletta, 2003).

Measurement Issues

The literature on measurement of change is substantial and extends over several decades (e.g., Harris, 1963). Several of the measurement issues that plagued investigators for so long were summarized and discussed by Bereiter (1963) as "persisting dilemmas in the measurement of change." These included (a) correcting for unreliability in the measurements being studied for evidence of change; (b) dealing with the fact that as the correlation increases between scores at two time points, the reliability of change (difference) scores decreases and vice versa; and (c) coping with the idea of the meaning of change scores when there is no objective physical dimension of change to which the former can be related (Cattell, 1966).

Cronbach and Furby (1970) captured a number of genuine problems associated with change scores. One consequence of their critique was the widespread reliance on indirect measurement of change via the assessment of stability. That is, longitudinal data were used to compute test–retest correlation coefficients that measured the stability of individual differences between two points in time, often referred to as *stability coefficients* (Bloom, 1964). Individual differences in change were identified when the stability coefficient was less than 1.0. There were several problems with such an approach. First, unreliability was confounded with instability, and it was impossible to know, without statistical corrections, the actual degree of stability in the variable. Second, the indirect nature of the stability coefficient and its indifference to the temporal lag between occasions of measurement made it difficult to scale rates of change in a meaningful way. Third, stability coefficients reflected change between two points in time, without attention to patterns of change over multiple observations.

In hindsight, avoidance of change scores due to the problem of low reliability was a misguided impediment to progress in developmental research (Rogosa, Brandt, & Zimowski, 1982). The critical issue in determining reliability of a change score is whether there are true individual differences in rates of change in the psychological construct. When the variance in true change is zero, one cannot reliably measure individual differences in change—because there aren't any. Change scores merely reflect the flux of random measurement error. When there is substantial variance in change, difference scores are reliable, valid, and interpretable. Other, more elegant ways of indexing change may be preferred (e.g., nonlinear regression slopes), depending on the nature of developmental change, but that does not invalidate the change score as a useful description of change between two points in time.

Design Issues

Just as measurement issues need to be confronted directly in gearing up to assess changes, so must one attend to a variety of design and modeling issues. Traditionally, measurement of intraindividual change requires repeated observation of the same individuals over time, as in (but not restricted to) longitudinal panel designs that have played a key role in aging research (Baltes & Nesselroade, 1979; Schaie & Hofer, 2001). To build a useful picture of changes, it matters a great deal, for example, which variables are measured, how often and over what intervals they are measured, and on whom the measurements are obtained.

For instance, the answer to the question *Who is measured?* should consider such problems as the confounding of the actual change with unwanted selection effects such as regression toward the mean (Campbell & Kenny, 1999; Nesselroade, Stigler, & Baltes, 1980). Regression due to measurement error produces an artifactual change that can be confused with true change in the underlying construct. If we elect to examine the effects of an intervention only on the people who "need it most" (i.e., the extreme scoring groups) we run the risk of totally confounding the effect of the intervention with regression toward the mean effects.

The answer to the question *What is measured?* also impacts directly on the level of generality of conclusions regarding change. Often, the variables that are actually measured are merely proxies for more theoretically interesting latent variables that cannot be measured directly. Which empirical variables are chosen, and how well they represent the latent variables of interest, is critical to the validity of a developmental design. Little, Lindenberger, and Nesselroade (1999) presented a systematic examination of variable selection on factor analysis results. Their principal conclusion was that the level of homogeneity in manifest variables (e.g., items) designed to measure a latent variable could limit validity of outcomes. In general, it is desirable to have broad nomothetic span (Embretson, 1983), that is, to select diverse measures of the construct so as to minimize the likelihood of obtaining factors that are specific to methods of measurement or reflect narrowly defined, specific sources of variance (Cook & Campbell, 1979). Depending on the number of indicators and their relationships to the latent variable, the optimal level of indicator correlations might be moderate rather than high. Furthermore, indicators validated with static assessments of construct validity may have unwittingly been designed so as to be insensitive to change and lability (Molenaar, Huizenga, & Nesselroade, 2003), as when scale items are selected for their high test–retest correlations. For example, trait measures with item stems or instructions that emphasize responding as one "generally feels" rather than "as one feels at this moment" are less sensitive to intraindividual variability over repeated measurements.

How often? and *Over what intervals?* the measurements are taken are also key research design considerations in assessing change. Measurement intervals that are too short or too long in relation to the nature of the phenomenon being studied can produce data that in some cases are overly sensitive to measurement errors and, in other cases, are insensitive to variability and change in the system (Boker & Nesselroade, 2002; Nesselroade & Boker, 1994; McArdle & Woodcock, 1997). In order to observe the dynamics of change, one must select observations temporally in a way that captures the changing states of variables as they change

(Nesselroade, 1991). For example, the timing of measurement occasions in a longitudinal panel design dictates whether one should model lagged effects (effects from $T - 1$ to T) or simultaneous effects (effects within time T) in autoregressive structural equation models of the data (Gollob & Reichardt, 1987; Kessler & Greenberg, 1981). For instance, if the interval between a putative cause and its putative effect is shorter than the interval between the repeated observations, simultaneous effects might well be favored over lagged effects. A further complication is that different assessment intervals may be optimal for different variables. Variables evincing circadian rhythms and those manifesting monthly cycles cannot be modeled optimally with the same repeated measurement intervals.

Studies of intraindividual variability and change (e.g., Cattell, Cattell, & Rhymer, 1947; Hultsch, MacDonald, & Dixon, 2002; Nesselroade, 1991) have led to the realization that describing changes in mean levels of variables is not always adequate for detecting key features of developmental change. Short-term change and variation in the organism's behavior also conveys important information regarding the phenomenon under investigation. For example, recent research on aging and cognition suggests that circadian rhythms can have marked impact on tasks requiring attention and executive control, thereby influencing estimates of age differences in cognitive performance (Hasher, Zacks, & May, 1999). Performance can also vary dramatically within a person over weeks, even when the time of day and day of week of the testing is held constant (Hampson, 1990; Hertzog, Dixon, & Hultsch, 1992). Reliable assessment of intraindividual variability requires multiple measures on the same individuals and requires design decisions about how to create and administer nonreactive tests and how often and on what schedule to assess individuals. Separating intraindividual variability from longer term intraindividual change requires repetitive measurement over relatively short time intervals (Nesselroade, 1991). This requirement has led to the research design proposal that the measurements be conducted in "bursts" rather than as single occasion activities. Here, again, research design is seen to impinge directly on the measurement of change.

Finally, the notion of "planned missingness" is also a design matter of considerable import for research on change (McArdle, 1994). The theme of the "planned missingness" perspective is that not all measures need be given to all participants at all occasions of measurement. Instead, measures may be selectively given to individuals in light of an overall plan that still allows estimation of all parameters of interest, provided this feature can be accommodated by the method of statistical analysis. Recent work by McArdle and Hamagami (1992, 2001), for example, has helped to formalize the methods and procedures needed to model change in such cases.

Modeling and Analysis Issues

A number of key decisions are involved in selecting and specifying a statistical model—most of which precede attempts to implement a particular statistical analysis with available data. For example, data sets often contain more variables than can be included in a given model, given the need for an adequate ratio of cases to model parameters (in order to yield stable estimates; see Bentler & Chou, 1987). Another key issue is how to handle nonnormal distributions. Although techniques are available to estimate and analyze ordinal rather than interval scales (e.g., by estimating polychoric correlations), many of these approaches ignore information about variances (Jöreskog, 1990). This approach can be problematic for the measurement of change, where diverging developmental trends lead to changes in variability over time, and where estimates of change from standardized variables can be misleading (e.g., Rogosa et al., 1982).

Furthermore, decisions on how to aggregate data across time (e.g., whether to ignore variation in the length of test–retest intervals in longitudinal panel data) or persons (e.g., whether to include or exclude specific types of cases from the sample) can have a profound impact on the validity and interpretation of modeling outcomes. For decades, the literature has harbored the occasional plaint that group statistics such as the mean represent every one and no one (e.g., Lamiell, 1997). This is a critical issue for some statistical models of developmental change. For instance, latent growth curve models (LGMs) represent individual differences in the shape of developmental functions as deviations from the average developmental function (McArdle, 1988; McArdle & Anderson, 1990). An important issue, then, is whether the aggregate developmental curve captures in a meaningful way the variation in the individual curves that the scientist is trying to study. Consider Figure 1, which we have borrowed from Baltes and Labouvie (1973). The underlying premise of the figure is that each individual is characterized by a two-limbed developmental function: (a) stability in function for a substantial period of time, followed by (b) late-life terminal decline (Berg, 1996; Bosworth, Schaie, & Willis, 1999; Riegel & Riegel, 1972). Because the age of onset of the

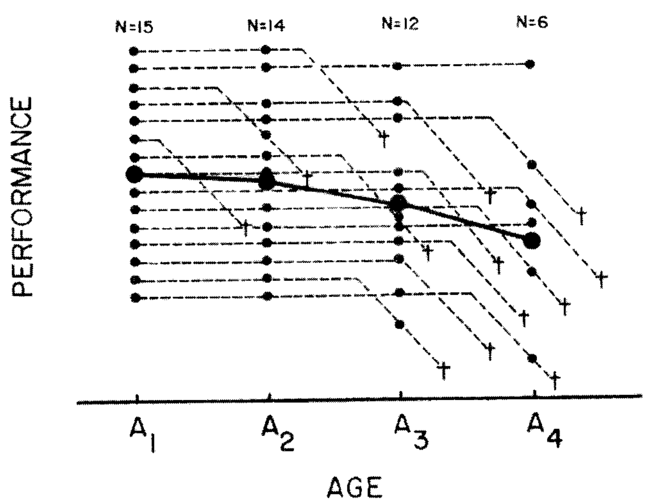

Figure 1. The effect of age-related increases in the frequency of terminal decline patterns on mean age functions. Note that the mean curve reflects the weighted combination of the invariant developmental function (scaled in time to death), with variable onset of decline and an increasing hazard of experiencing terminal decline with advancing age. From "Adult Development of Intellectual Performance: Description, Explanation, and Modification," by P. B. Baltes and G. V. Labouvie, in *The Psychology of Adult Development and Aging* (p. 174), edited by C. Eisdorfer and M. P. Lawton, 1973, Washington, DC: American Psychological Association. Copyright 1973 by the American Psychological Association. Reprinted with permission of the author.

terminal decline is a random variable, the average developmental function, shown in Figure 1, is a curvilinear trend that indirectly reflects the increasing probability of terminal decline as a function of increasing age. Although the influences on this aggregate curve behave lawfully, the curve itself accurately describes the shape of none of the individual developmental functions in the sample. This problem is well-known in statistical models for learning, development, and growth (Estes, 1956; Wohlwill, 1973) as a type of aggregation bias. It is a critical issue in the selection of a statistical model for describing developmental change, although it has actually received relatively little attention in the literature describing newer modeling approaches to measuring change.

Of course, the essence of scientific inference is generalization that will require some forms of data aggregation. Techniques like latent growth curve analysis, and other available tools for modeling change-related phenomena, have advanced our science because they provide a beneficial and appropriate means to address many vexing issues regarding measurement of change. The well-known difficulties of dealing directly with raw difference scores, given issues of reliability (errors of measurement) and validity, can be surmounted by structuring and measuring growth and decline indirectly with structural equation models (McArdle, 1988; McArdle & Nesselroade, 1994; Raykov, 1999). Because these statistical models allow explicit specification of how errors of measurement influence the means and covariance matrices of variables, these models can, in theory, separate change in latent variables from errors of measurement and address the unreliability—invalidity dilemma (Bereiter, 1963). However, such models can best achieve the goal of valid measurement in an integrated approach that selects measures and times of observation with a statistical modeling approach firmly in mind.

Tests of Measurement Equivalence and Construct Invariance

A critical goal for researchers is to create measures enabling valid assessment of qualitative changes in constructs and justifying the inference of quantitative changes in psychological constructs (Baltes & Nesselroade, 1970). Measurement equivalence, indicating that a measure of a psychological construct has equivalent measurement properties at different ages or times, is a necessary condition for treating differences in those units of measurement as reflecting quantitative differences in a given construct (Baltes et al., 1988; Labouvie, 1980).

Structural modeling techniques have been widely used to model the empirical behavior of multiple self-report scales in gerontological research (Alwin, 1988; Schaie & Hertzog, 1985). These techniques have been widely used to evaluate age differences in the factor structure of self-report items (e.g., Hertzog, Hultsch, & Dixon, 1989; Hertzog, Van Alstine, Usala, Hultsch, & Dixon, 1990; Liang, 1985; Zautra, Guarnaccia, & Reich, 1988). This approach can also be extended to longitudinal observations on a set of items to evaluate the invariance of these measurement properties over time and to separate the consistency, or reliability, of the scales from the stability of individual differences in the true scores (e.g., Alwin & Krosnick, 1989; Hertzog & Nesselroade, 1987). In longitudinal studies, a valid concern is that the act of measurement itself will alter the measurement properties of the scales—individuals will react to being measured in ways that alter their responses on the scales. Alternatively, developmental change may cause shifts in the ways that questions or tasks are interpreted. For example, changes in self-rated depression could reflect changes in actual depression or merely changes in how individuals interpret and respond to the scale's items.

Studies evaluating longitudinal measurement equivalence have typically found relative time-related invariance of scale properties in adult samples (e.g., Hertzog & Nesselroade, 1987), but such issues should be evaluated separately and anew for different measures and/or different populations. Maitland, Dixon, Hultsch, and Hertzog (2001) rejected the hypothesis of temporal invariance in scales from the Bradburn Affect Balance Scale (Bradburn, 1969) in an older population; one item shifted the magnitude of its loading on the Negative Affect factor over time. Given the longitudinal invariance of other item factor loadings, they recommended use of latent variable analysis to evaluate age changes in mean negative affect with this instrument. Given the lack of complete metric invariance, mean scale scores would be influenced both by level of negative affect and shifts in measurement properties of items over time.

Tests of factorial invariance are not just important tools for evaluating measurement properties of tests and scales. Evaluation of developmental changes in the organization of psychological constructs also relies on comparative factor analysis of an appropriate set of indicators (Reinert, 1970). For example, one important hypothesis about the multidimensional construct of intelligence (Carroll, 1993) is that basic ability factors, such as inductive reasoning, spatial relations, or verbal comprehension, become less differentiated in old age. Here, too, measurement equivalence issues play a role, because changes in the construct validity of measures can also alter the factor structure. Changing constructs must be distinguished from changing construct validity.

Hence, the critical questions are the following: Does the same construct operate in the same way at different points in the life span, and if so, are the hypothesized measures of that construct equally valid (Meredith & Horn, 2001; Schaie & Hertzog, 1985)? To evaluate these questions, one requires a sampling design on both persons and measures in a temporal context that enables comparisons across psychologically meaningful portions of the adult lifespan. Selection of measures should be achieved through a priori identification of relevant target constructs of interest and a set of measures that, according to hypothesis, are determined by these latent constructs. Ideally, a core set of variables should be selected to conform to simple structure (i.e., one variable loading on only one factor; Thurstone, 1947), but this can be difficult to achieve in practice and should be viewed as a desideratum rather than a requirement.

Owing in large part to work by Meredith and colleagues (e.g., Meredith, 1993; Meredith & Horn, 2001), much is already known about what should be expected regarding the invariance of factor structure of a set of variables. Even if a common factor model is the correct model for an entire population of interest, a partition of subgroups from that population will not necessarily produce equivalent factor structures in each subpopulation (Little et al., 1999). Selection, when modeled as a regression filter by selection variables on the whole population, produces subpopulations that differ in means of the latent variables (i.e., means of the implied factor scores) and also produces differing factor variances, factor covariances, factor means, unique (residual) means, and unique variances

(Meredith, 1993). Clearly, many variables differ as a function of chronological age, so one cannot generally expect that the factor structures for different age groups will be completely invariant, even if the same underlying model aptly captures the structure of the variables in the aggregate population. However, the general type of selection studied by Meredith (1993) will still result in an invariant factor pattern matrix (factor loadings), provided that these loadings are estimated in raw-score metric (rather than in the usual standard-score form). In effect, the cosines of the angles of the dimensional space defining relations among variables remains invariant, but selection alters the length of the vector needed to span the observed variances defined by these dimensions. Hence, tests of measurement equivalence using empirical data require confirmatory analysis of covariance matrices (and, perhaps, estimation of factor and unique variable means as well; Meredith, 1993).

Some authors have claimed that configural invariance—in which different age groups have the same pattern of loadings of variables on factors but in which the numerical values of these loadings vary—is more likely to be the rule, and metric invariance the exception (Horn & McArdle, 1992). Actually, metric invariance of ability factor structure across adult age groups has been found in a number of studies (e.g., Brickley, Keith, & Wolfle, 1995; Hultsch, Hertzog, Dixon, & Small, 1998; Stricker & Rock, 1987). Generally speaking, however, configural invariance of ability factor structures has been found, even when metric invariance has not (e.g., Schaie, Maitland, Willis, & Intrieri, 1998; Zelinski & Lewis, 2003). Candidate explanations for a lack of metric invariance include age-related changes in the structure of the constructs or age-related changes in measurement properties of the variables (Schaie & Hertzog, 1985). The latter explanation could involve contamination of the common factor space by specific sources of variance, including method variance, that are not the target of the selected set of measures (Meredith & Horn, 2001). Specific or unique components can covary, even though the factor model assumes they are orthogonal to each other and the common factors specified by the model. Under such circumstances, developmental change in some specific components of variance, but not others, may lead to factor loading estimates that vary between age groups (Hertzog & Bleckley, 2001).

Similar issues apply to the invariance of factors over time in longitudinal data (e.g., Tisak & Meredith, 1990). The available evidence suggests that abilities demonstrate unchanging configurations of factors (i.e., the basic pattern of relations of variables to factors is unchanging). It also appears that metric invariance may hold, with equivalent unstandardized factor pattern weights (Hertzog & Schaie, 1986; Hultsch et al., 1998). Aging appears to alter the correlations of cognitive ability factors in some but not all data sets (cf. Zelinski & Lewis, 2003).

Metric invariance was not obtained in one longitudinal study evaluating the effects of ability training on intellectual factor structure (Schaie, Willis, Hertzog, & Schulenberg, 1987). Instead, partial metric invariance was found (see Byrne, Shavelson, & Muthen, 1989), with the reasoning test used for training purposes showing changes in its loading on an Inductive Reasoning factor. This outcome suggested that training on a particular set of induction test items alters the relationship of the test using that item type with other tests of inductive reasoning.

Seemingly small variations in the selections of measures can have an important impact on longitudinal models of individual differences in change. Hertzog, Dixon, Hultsch, and MacDonald (2003) showed that merely substituting one measure of working memory for another in a latent variable model of change produced substantial shifts in the estimated relationships between changes in working memory and other cognitive constructs, despite the fact that the set of working memory measures in question all showed substantial correlations with one another. They did not, however, form a stable and coherent latent variable, given divergent relationships of working memory measures to other constructs. Explicit evaluation of the factor structure of measures is a necessary precursor to valid structural regression analysis.

There are a number of important current issues involving assessment of factorial invariance. One potentially controversial issue is whether assessment of invariance requires simultaneous analysis of mean and covariance structures (Meredith, 1993). Factor analysis is predicated on a model of multiple components of variance that determine test score performance, and such components can be divided into common factor variance and sources of test-specific variance. The distinction is somewhat arbitrary, in that a specific source of variance can be made a common source of variance by virtue of selection of variables to mark factors (Little et al., 1999). Nevertheless, the distinction is valid for any given set of factors and measures defined by a sampling plan on variables. Meredith and Horn (2001) argued that age group differences in specific factor means could bias tests of factorial invariance, unless these factor means are explicitly modeled.

Although we respect the argument concerning analysis of means in studies of factorial invariance, we believe its logic is grounded in a static, psychometric perspective on measurement and tests of measurement equivalence. As developmentalists interested in the study of change and process, we do not believe that the processes that account for covariance structures are necessarily going to be the same as the processes that determine mean levels of the component variables (see later section *Reconceptualizing Measurement "Error"*). Therefore, we would not always expect a single, restricted measurement model to account adequately for both means and covariance structures. Consider, for example, generational differences in mean level of physical stature that coincide with no detectable generational differences in the covariation pattern of body parts. Something (diet?) can universally increase or decrease physical stature over time or between generations without affecting how the sizes of different body parts covary with each other. We advocate renewed attention to the issue of whether simultaneous modeling means and covariance structures ought to be considered a necessary feature of tests of factorial invariance.

Statistical Models for Measuring Intraindividual Change: Recent Developments

Autoregressive Structural Equation Models

When structural equation modeling software first became widely available in the 1970s, the predominant method for modeling change in longitudinal panel data was use of an autoregressive model (e.g., Alwin, 1988; Dwyer, 1983; Jöreskog, 1979; McArdle & Aber, 1990; Rogosa, 1979; Schaie & Hertzog, 1985).

Variants of the autoregressive approach are still strongly recommended today (e.g., Rudinger & Reitz, 2001). In this approach, occasion-specific latent variables are modeled at each time of measurement, as is a flow of influence from earlier to later points in time. Figure 2 depicts a first-order autoregressive model, in which a single latent variable L at time T is modeled as being determined by the same latent variable at time $T - 1$. This regression coefficient is the unstandardized equivalent of the simple stability coefficient already mentioned, because it estimates the stability in rank orders of individuals between the two points in time. Instability is reflected in the residual variance of the disturbance term, d, at time T, controlling for autoregression. Thus the residual variance is an indirect measure of individual differences in change from times $T - 1$ to T. Given two (or more) latent variables, measured at two (or more) points in time, cross-lagged or simultaneous regression models can be specified that start with an autoregressive model for each latent variable. Variables that predict a target variable at time T, controlling for autoregression, are interpreted as causes of change from time $T - 1$ to T (Kessler & Greenberg, 1981). This class of models has been fairly widely used in gerontological research (e.g., Finch & Zautra, 1992; Hertzog, Cooper & Fisk, 1996; Newsom, Mishishiba, Morgan, & Rook, 2003; Schooler, Mulatu, & Oates, 1999).

An interesting potential application of autoregressive models is evaluation of direction of causation, either through the comparison of magnitudes of cross-lagged regression coefficients (Rogosa, 1980) or simultaneous, reciprocal regression effects (e.g., Schooler et al., 1999). Whether one would specify lagged or simultaneous effects is in part a function of the frequency of longitudinal sampling, relative to the dynamics of the causal process (Kessler & Greenberg, 1981), as was alluded to in the earlier discussion of timing of measurement. Whether comparisons of relative magnitudes of standardized regression coefficients can identify relative importance of two or more variables in a causal process is a more difficult question, one that is not fully resolved (Dwyer, 1983; Gollob & Reichardt, 1987; Rogosa, 1979).

There is little question that the use of autoregressive structural equation models for this purpose is superior to ordinary cross-lagged correlation (Kenny, 1979) or multiple regression analysis, in large part because specific error structures and regression relationships between residual (or disturbance) terms can be explicitly modeled (e.g., Kessler & Greenberg, 1981). Hence, estimates of stability and mutual influence between variables are corrected for random measurement error, and sources of systematic error can be specified and modeled, as well.

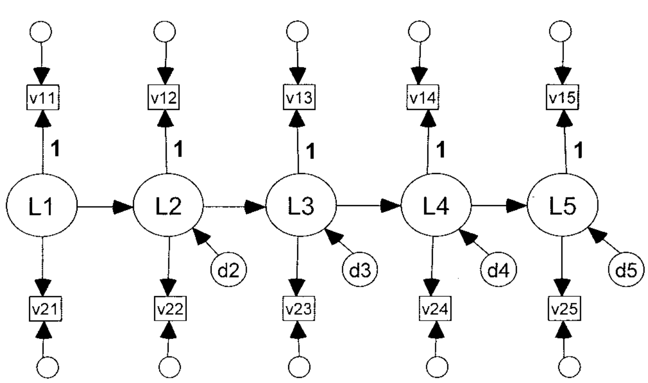

Figure 2. A hypothetical autoregressive model with a latent variable L_t ($t = 1 \ldots 5$), a first-order autoregressive process (e.g., $L_1 \rightarrow L_2$), and two observed variables, v_{1t} and v_{2t}, measuring L_t at each occasion of measurement. The metric of L_t is defined by the fixed-1 loading of v_{1t} on L_t. Stability is indirectly reflected in the autoregressive coefficients, and instability is indirectly reflected in the variance of the disturbance terms, d_t ($t = 2 \ldots 5$).

Because the autoregressive model has been widely used and evaluated, much is known about its strengths and weaknesses. The model is potentially useful for identifying flows of influence in panel data. It has the additional virtue that stability of individual differences in underlying constructs can be estimated from the standardized autoregressive coefficient even if there is a lack of full measurement equivalence in the measures over time, as long as one can assume that the measures maintain configural invariance over time (i.e., they continue to measure the same latent variable). There are several potential limitations of the autoregressive approach, including the following: (a) the autoregressive model parameters do not explicitly model change, but rather indirect manifestations of change; (b) comparison of cross-lagged regression coefficients does not necessarily identify leading and lagging (cause and effect) relationships in a panel design (Rogosa, 1980); (c) the autoregressive model carries implicit assumptions of entropy (decreases) in latent variable correlations that can be problematic (Hertzog & Nesselroade, 1987); (d) dynamic equilibrium in causal structure may not be adequately reflected in parameters estimated from panel data, in which a first point of measurement is generally not the point of inception of the equilibrium process (Curran & Bollen, 2001; Dwyer, 1983; Gollob & Reichardt, 1987); (e) the autoregressive parameters are indifferent to the functional form of change over time (e.g., exponential rates of learning) and to individual differences in the parameters of any functions reflecting factors that govern that change; and (f) change in a variable cannot, in itself, be modeled as a cause of change in other variables, even when change is the theoretically meaningful causal concept.

Latent Change Models

McArdle and Nesselroade (1994) introduced an alternative approach, latent change models. These models still specify occasion-specific factors, but they do not use an autoregressive approach to estimate stability coefficients. Difference scores are not calculated and are not directly factored. Instead, additional higher order latent variables are specified that measure latent initial level and change for each latent variable. Figure 3 illustrates this type of model. Such a model can be executed on a longitudinal covariance matrix, with or without adding the vector of observed means. The chief advantage of this model is that changes in the latent variables are represented as factors. Thus, it is possible to estimate the variance of the latent change factor as a parameter. Rejecting the null hypothesis of zero variance indicates that there are reliable individual differences in change. Such a test is relevant to claims that individual differences in latent variables are highly if not perfectly stable (e.g., Costa & McCrae, 1994). Small, Hertzog, Hultsch, and Dixon (2003) used latent change models to demonstrate that, despite impressive longitudinal stability of the so-called Big 5 personality factors, reliable individual differences in adult personality change could be detected over a 6-year interval. Latent

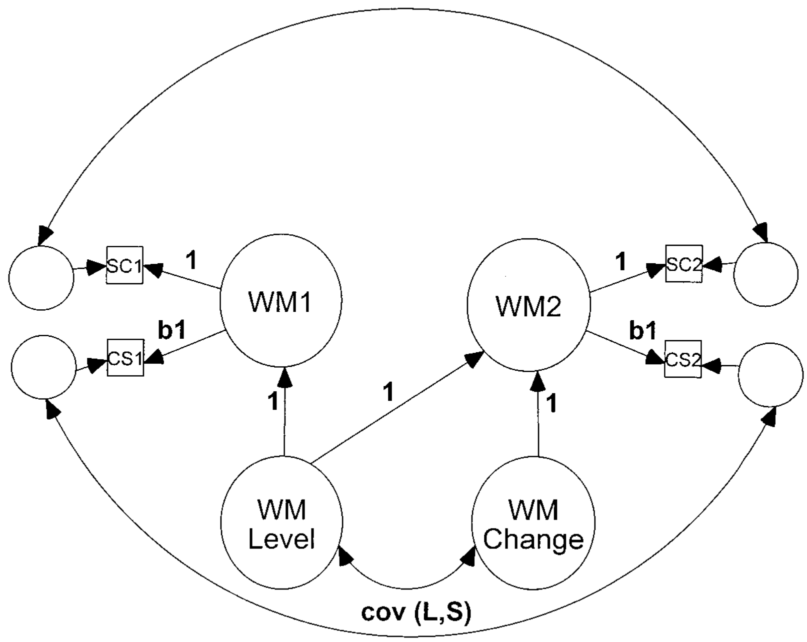

Figure 3. Latent change model for two measures of a latent working memory variable (WM), measured at two points in time (Occasions 1 and 2). The fixed-1 regression coefficients involving the latent variables implicitly define the latent variable of Working Memory Level as equal to Working Memory at Wave 1 (WM1) and Working Memory Change latent variable as the difference between Working Memory at the two waves (WM2 − WM1). Residuals of the two measures of working memory (Sentence Construction and Computation Span) are allowed to covary across occasions, reflecting stability in systematic errors over time.

change models can also be used in structural regression models in which change in one variable predicts change in another variable. Hultsch et al. (1998) and Hertzog et al. (2003) used latent change models to evaluate whether change in processing resource variables (speed and working memory) are associated with changes in more complex cognitive variables, such as inductive reasoning and episodic memory. Hultsch, Hertzog, Small, and Dixon (1999) found that changes in self-reported health did not predict changes in cognitive constructs like working memory but that changes in intellectually engaging activities did predict changes in cognition.

A limitation of the latent change models as just illustrated is that they are restricted to two occasions of measurement. As such, they implicitly assume linear change between the two time points, which may not be an apt description of change over longer time intervals. However, the models can be extended to multiple successive-difference or change factors across three or more occasions (see McArdle & Aber, 1990).

LGMs

Basic model specification. The latent change models just described can be considered a special case of LGM (Duncan & Duncan, 1995; McArdle & Anderson, 1990). LGMs generally specify a Level and a Shape latent variable for each construct. Usually (but not always), the Level variable represents the initial point of measurement in a longitudinal data set. The Shape variable captures individual differences in rates of change over time. Means of the observed variables are typically also modeled as being determined by the means of the latent Level and Shape. The functional form of change, which we refer to as the *basis function* for scaling change, is defined by values of the loadings of observed variables on the Shape factor. This basis for Shape can be defined with a variety of metrics (e.g., McArdle & Bell, 2000; Rovine & Molenaar, 2001), including a nonlinear basis function (e.g., a Gompertz function) defined a priori (Browne & duToit, 1991). A nonlinear basis function can also be estimated from the data, which is a powerful alternative to polynomial decomposition of the curve (McArdle, 1988; McArdle & Bell, 2000). LGMs can be specified for a single variable or multiple variables so that correlations of changes between variables can be evaluated. The technique is even more powerful when applied to multiple latent constructs, so that (a) latent occasion-specific factors are specified and (b) Level and Shape factors are specified that relate to these factors, as in the latent change models described previously. The chief advantage of this approach is optimal correction for measurement error, relative to models estimating curves for a single measure of a construct (McArdle, 1988; Sayer & Cumsille, 2001).

Figure 4 depicts an LGM for a single observed variable, measured at five occasions (*v*1 . . . *v*5). At each occasion, the variable is modeled as having components due to initial level, change from initial level, and measurement error (the d_t in Figure 4). The Level factors are defined by fixed-1 loadings of all variables for a given latent construct on its Level factor. In Figure 4, the scaling of the basis function for both variables implies linear growth or decline. This is reflected in the equal-interval increases in the fixed factor loadings on the Shape latent variable over occasions of measurement. In the example given, the scale of the basis vector is set as

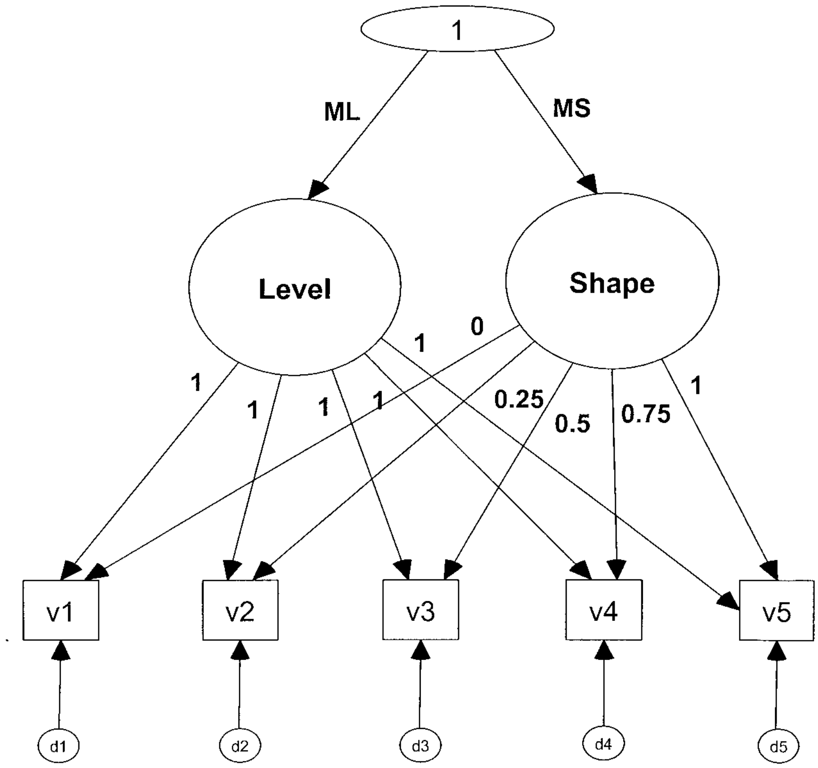

Figure 4. A univariate latent growth curve model for a single variable measured at 5 occasions (*v*1 through *v*5) in a panel study, using the Level, Shape, and disturbance term (*d*1 . . . *d*5) latent variables. The oval defines the constant 1, which is used to estimate the mean level (ML) and mean shape (MS) of the developmental function. Variances in the latent Level and Shape variables reflect individual differences in the shape of the developmental function defined by the regression weights of observed variables on the Shape latent variable. The fixed weights shown imply a linear growth function (given equal time intervals); the weight of .5 for the loading of *v*3 on Shape indicates that 50% of the total change from Time 1 to Time 5 has occurred by Time 3. The model as drawn assumes the covariance of Level and Shape is 0; a covariance parameter between Level and Shape can be added to the model.

proportion of change between initial status (fixed 0) and final status (.00, .25, .50, .75, and 1.00—see Rovine and Molenaar, 2000, for an alternative method of scaling the basis function with integer weights). The fixed-0 and fixed-1 parameters at the endpoints are needed to define the Shape factor and to determine its metric. If the Shape factor weights for Occasions 2–4 were estimated from the data, a nonlinear shape could be implied. For example, estimated weights of .0, .1, .2, .4, and 1.0 would imply change that is accelerating in rate late in the observational time period. This basis function applies equally well to growth and decline (individuals declining over time would, implicitly, have negative scores on the Shape factor). However, a critical assumption of the model is that all individuals in the population have a growth (or decline) process that is governed by this shape function. By assumption, individuals can vary in the amount and rate of developmental change, not in the functional form of development.

Important parameters in the model include means of the Level and Shape factors (initial level and average change over time), variance of Level and Shape variables (measuring initial interindividual variability and individual differences in latent change, respectively), covariances of Level and Shape variables within a construct (indicating whether magnitude of change is dependent on initial level), covariances of Level and Shape parameters between constructs (indicating initial variable relations and covariances in change between variables over time), and error variances over time. These latter parameters allow for deviations of the observed scores from the pattern of growth defined by the Shape parameter. Restricted hypotheses can be tested by likelihood-ratio tests (e.g., testing for individual differences in change by evaluating whether the variance in Shape is reliably greater than zero).

LGM can be extended to multiple observed variables and to include a growth process on latent variables rather than observed variables (McArdle, 1988). The critical issue for the analysis of change is often the covariance between Shape parameters for different variables, because these covariances address whether individual differences in rates of change are correlated for different variables. There is some controversy about whether the covariance of Level and Shape for a given variable is actually interpretable, substantively, in terms of individual differences in change (see Rudinger & Reitz, 2001). In part, this is because the covariance of Level and Shape is not invariant with respect to choice of scale (i.e., how time is centered by the fixed zero loading on Shape; see also Rovine & Molenaar, 1999).

Alternative LGMs for sequential data. A major issue in applications of growth curve models to sequential data is that, unlike a traditional single-cohort longitudinal design, individuals will often vary substantially in age (and birth cohort membership) at the inception of the sample (see Schaie, 1996). This heterogeneous sample will then be followed longitudinally, often at relatively constant retest intervals. This kind of longitudinal sequence (Baltes et al., 1988) generates relatively wide spans of chronological age, cross-sectionally, and generates intraindividual change data slowly, as years pass.

These kinds of designs immediately raise the question of whether the growth curve model should use time—or occasion of measurement—versus age, as the way of defining the basis curve representing change (McArdle & Bell, 2000; Mehta & West, 2000). Figure 4 implies that occasion of measurement in a panel study is used as a basis for organizing growth. This specification would be appropriate if the form of change depends on time or historical period. However, organizing growth curves in this way, one may encounter the problem that the functional form of change is governed by an age-graded developmental process, not by the mere passage of time. If, in fact, the developmental curve is entrained by processes associated with chronological age (as in ontogenesis) and if the functional form of the basis curve is nonlinear across age levels, then a standard LGM produces biased estimates of the average change function and individual differences in the shapes of that function (Mehta & West, 2000). Under these conditions, more complex models are required that use an age-graded basis function. The LGM model can be expanded to define multiple subgroups. Each member of a group would have the same age span covered in its slice of the longitudinal sequence; groups would differ in their realized age spans (McArdle & Hamagami, 1992; Mehta & West, 2000). Estimation of the aggregate developmental function in typical LGM invokes what is called the *convergence assumption* (i.e., equivalent expected value of cross-sectional and longitudinal sources of age effects; McArdle & Bell, 2000). Indirectly, this assumption involves assuming no cohort differences and negligible impact of other confounds (e.g., attrition, practice effects) on longitudinal and cross-sectional estimates of age change.

Convergence assumptions can be problematic. Rates of attrition are often quite substantial in longitudinal studies (Hultsch et al., 1998; Schaie, 1996), and attrition is often related to the developmental construct of interest. For example, depressed individuals may be less likely to volunteer to be assessed again, and individuals who have cognitive impairments may be more likely to die (e.g., Berg, 1996). A common way of approaching this problem is to analyze data only for those who return across all available occasions of measurement in the longitudinal study, even though this introduces bias to the estimated level and change functions and may constrain external validity of longitudinal results to the larger population. That is, one obtains an estimate of average level, average intraindividual change, and individual differences in change for a residual population that is positively selected relative to the inception sample and to the population at large.

McArdle and Hamagami (1992, 2001) have argued instead that all available data should be used to estimate the developmental function. These authors have introduced complex multiple-group LGMs that include all individuals, irrespective of their participation history. One specifies groups of persons who are homogeneous with respect to patterns of missing data. For example, in a four-occasion longitudinal panel study, one group might have all four occasions, and another group might have data on only the first occasion, another group might have data on the first, third, and fourth occasions of measurement. Obviously, as the number of occasions grows, the number of possible combinations of missing data patterns (and hence, groups) grows as well. Then, one estimates the model constraining the key parameters (e.g., means of Level and Shape, covariances of Shape parameters between variables) to be equal across the groups.

Extensions of growth curve models to more complex sampling designs are possible and have considerable potential benefit. A key problem in longitudinal studies is that practice effects or reactive effects of testing are confounded with occasion of measurement. Practice effects can bias estimates of average age changes, and individual differences in practice benefit can bias estimates of individual differences in rates of developmental change (Donaldson & Horn, 1992; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001; Salthouse, 2000; Schaie, 1977). McArdle and Woodcock (1997) considered how to extend LGMs for longitudinal sequences to incorporate estimates of practice effects.

To some degree, the field is only beginning to explore what is possible within different classes of latent variable models and what can be gained by combining features from different classes. For example, Curran and Bollen (2001) argued for a combination of autoregressive models and LGM into a hybrid that preserves important features of both kinds of change models. Rudinger and Reitz (2001) argued that their reparameterization of autoregressive models, allowing for autoregression from unobserved values of variables prior to commencement of observation in a panel design (see also Gollob & Reichardt, 1987), is superior to LGM. Undoubtedly, the next decade will see additional explorations of alternative specifications of change and the consequences these specifications have for measurement of development.

Applications of LGM to aging research. At present, there have been relatively few applications of LGMs in empirical research in gerontology, although the rate of application appears to be increasing. McArdle, Hamagami, Elias, and Robbins (1991) applied LGM with incomplete data to the study of relationships between hypertension and cognition. There were some differences between hypertensive and normotensive groups in patterns of cognitive change. Lynch and George (2002) demonstrated that longitudinal changes in late-life depressive symptoms were associated with increases in psychological stress associated with loss-related events (e.g., widowhood). Their data support the claim that a substantial part of depressive symptoms in older adults are reactions to stressors. Hofer et al. (2002) evaluated whether the apoE genotype was associated with differential cognitive change in a longitudinal study using LGM. Anstey, Hofer, and Luszcz (2003) used LGM on latent variables to evaluate whether sensory changes are as highly related to cognitive changes in longitudinal data as expected from cross-sectional data showing strong sensory-cognitive correlations. They found that these changes are not highly related, suggesting that accounting for cross-sectional age-related variance (Salthouse, 2000) is not necessarily identifying variance in age-related change. Christensen et al. (2001) evaluated whether longitudinal age changes were smaller in better-educated older adults and concluded that they were not. Latent cognitive slopes did not vary by educational level. McAuley et al. (1999)

used LGM to evaluate an intervention trial (with a 1-year follow-up) examining exercise regimen effects on self-efficacy and frequency of exercise participation. Individual differences in change in exercise efficacy were predicted by changes in exercise frequency and change in physical fitness (in the stretching–toning intervention).

In all these cases, modeling change in latent variables, correcting for measurement error, was critical to be able to evaluate relationships between variables over time. Furthermore, LGM permitted an evaluation of how changes in psychological variables predict changes in other variables, hypotheses that are not easily tested in autoregressive models for panel data.

Multilevel (Random-Effects) Models

Basic features. Random-effects models represent a relatively recent statistical approach that has been successfully applied to longitudinal data analysis (Bryk & Raudenbush, 1987; Goldstein, 1995; Raudenbush, 2001). The approach uses a particular form of the generalized linear model to specify fixed effects and random effects, and can handle nested MLMs. Hence we refer to these approaches as MLMs. Individuals (persons) represent one class of random effects, and so individual differences in level and change of psychological variables are treated as the lowest level in a possible hierarchy of interrelated variables. When applied to developmental data, the basic setup for such models is as follows.

Consider a longitudinal sequence, with persons having varying initial ages and age spans for which they are measured longitudinally. The simplest MLM for data from such a sequence would involve a single variable (e.g., fluid intelligence) to be modeled as a function of development, but multiple variables can be included. There are (at least) two basic equations to consider. The first models each individual's pattern of development in fluid intelligence as a function of age or time (e.g., intercept and slope parameters in straight-line growth). The second-level equation models individual differences in these regression coefficients as a function of individual differences in other variables. In this way, covariates or predictors of change can be incorporated into the model. Age, then, is treated as a random variable that differs between persons (at the start of sampling) and increases longitudinally within a person. Because the within-person regression coefficients specifically use age as an index variable, age trends in the data can be estimated (e.g., means at each age, individual differences in age changes).

MLM for change in a single variable produces estimates of average initial level and change, as well as individual differences in level and change, that are formally equivalent to the estimates obtained from an LGM (MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Raudenbush, 2001; Rovine & Molenaar, 2000, 2001). For univariate growth curves, then, the approaches are interchangeable, although there has been some debate about whether LGM or MLM is superior in certain applications. One of the key benefits to MLM for this purpose is the flexible handling of missing data, as well as the seamless treatment of convergence assumptions for age effects within and between persons (McArdle & Hamagami, 2001). Arguably, MLM is simpler than LGM when analyzing longitudinal sequences under convergence assumptions. Unlike LGM, MLM requires no grouping by homogeneous age and missing data patterns. Furthermore, MLM can be used when patterns of longitudinal sampling are irregular or even random across persons. LGM uses a group aggregation approach, which allows the exact spacing of longitudinal occasions for a given person to deviate from the aggregate spacing implied by group membership. This can introduce systematic bias in LGM estimates that is rarely modeled explicitly. Conversely, MLMs are typically limited to modeling growth and predictors of growth by using observed variables, not latent variables. Hence the type of growth curve on latent variables considered by McArdle (1988), Sayer and Cumsille (2001), and others cannot easily be estimated by using standard MLM techniques.

A major advantage of MLM is that the occasion-specific structure of the data is, in essence, ignored (but see Rudinger & Reitz, 2001). Because time (or age) is a critical index of reference in the equations, the regression estimates for level and slope, under missing-at-random assumptions, define the aggregate developmental function. For nonlinear developmental functions, more complex basis functions can be applied (e.g., polynomials) as in LGM. More critically, time or age itself is treated as a random variable, so the model is ideal for analyzing data where dates (and hence, ages) of sampling vary randomly between participants.

More recent developments include methods for formally specifying nonlinear basis functions (e.g., SAS PROC NLMIXED; SAS Insititute, 2000) and a new modeling approach, the bivariate dual-change score model (Hamagami & McArdle, 2001; McArdle & Hamagami, 2001) instantiated within an MLM. This model is also a hybrid between LGM and the cross-lagged panel autoregressive model discussed earlier (see Ghisletta & Lindenberger, 2003).

Applications. A number of studies have estimated intraindividual change scores or slopes for multiple-occasion data, using simple change scores or ordinary least squares (OLS) regression to estimate individual slopes and intercepts (e.g., Alder, Adam, & Arenberg, 1990; Christensen et al., 1999). These studies consistently find reliable individual differences in magnitudes of cognitive change during adulthood. However, a recent comparative analysis suggests that MLM approaches are generally superior to the analysis of change scores or other indirect methods for assessing predictors of change (Reynolds, Gatz, & Pedersen, 2002).

A series of articles used longitudinal data from the Boston Normative Aging Study to evaluate change in personality, health, and mental health (e.g., Aldwin, Spiro, Levenson, & Bosse, 1989; Aldwin, Spiro, Levenson, & Cupertino, 2001; Spiro, Aldwin, Levenson, & Bosse, 1990). This study is best served by an MLM approach to measuring individual change because of the irregular sampling intervals for study participants. Their general approach has been a two-stage analysis, in which the first step estimates slopes and intercepts for each individual, and the second step analyzes these parameter estimates with other predictors in regression models. Early work in this series also used OLS regression to estimate individual change slopes (e.g., Spiro et al., 1990); later work in this series has used SAS PROC MIXED (SAS Institute, 2000) to estimate parameters of latent curves (Aldwin et al., 2001). In general, this research supports the argument that adult changes in physical health, mental health, and personality covary.

Wilson et al. (2002) also used a two-stage estimation approach. They first used SAS PROC MIXED to estimate linear change slopes for individuals in their panel study and then conducted additional analyses on the slopes. Substantial correlations of slopes across different cognitive measures were found, which Wilson et al. argued was evidence for a higher order factor of cognitive change (based in part on a principal-components analysis of the estimated slopes). These results were relatively consistent with work that has used latent change models to examine the structure of cognitive change in late life (Hultsch et al., 1998), despite the fact that the analysis of the structure of change was not conducted simultaneously at the latent variable level.

Other research with MLMs has used simultaneous estimation of change parameters and covariate–predictor relationships. Weatherell, Reynolds, Gatz, and Pedersen (2002) showed that state anxiety was related to level of cognitive performance in a longitudinal study, but that neuroticism was not a risk factor for greater rates of cognitive decline. Lane and Zelinski (2003) demonstrated weak relationships between memory complaints and longitudinal changes in memory in a sequential sample, suggesting only limited predictive validity of complaints for actual cognitive changes. Kemper, Thompson, and Marquis (2001) showed that age was associated with increased rates of longitudinal decline in grammatical complexity and other measures of language production and that diagnosed Alzheimer's disease was associated with a accelerated rate of decline. Sliwinski and Buschke (1999) used MLMs to demonstrate that processing speed was an important covariate for predicting cross-sectional age differences in episodic memory, but that longitudinal changes in processing speed were not as important for predicting longitudinal changes in episodic memory. In the latter study, the modest relationship between speed and cognitive change would not have been anticipated from existing cross-sectional studies (e.g., Salthouse, 1996). Recent MLM longitudinal studies by Zimprich (2002) and MacDonald, Hultsch, Strauss, and Dixon (2003) have come to a similar conclusion.

Rabbitt et al. (2001) applied MLM to separate practice effects, mortality, and normative aging influences on a measure of general intelligence, using data from a large community sample of older adults. Practice effects were substantial, even in individuals who had relatively long retest intervals; adjusting for practice effects revealed a quadratic age-related cognitive decline function, with reliable decline detected after age 70.

McArdle, Ferrer-Caja, and Hamagami (2002) used nonlinear MLM to characterize changes in multiple intelligence factors, including fluid and crystallized intelligence, across the entire human life span. Exponential growth and decline curves were fit to the different abilities. They provided convincing evidence for the differences in shapes of the age functions, with steeper changes after maturity for fluid intelligence and general speediness than for crystallized intelligence. Finkel, Reynolds, McArdle, Gatz, and Pedersen (2003) used MLM to demonstrate accelerating cognitive decline after age 65 on multiple measures in the Swedish Adoption/Twin Study of Aging. Reynolds, Finkel, Gatz, and Pedersen (2002) analyzed the same data set to show that slopes of change showed less heritability than levels of cognitive performance, and they attributed differential decline amongst members of twin pairs to environmental causes. Ghisletta and Lindenberger (2003) used the bivariate dual change score model to evaluate changes in processing speed and semantic knowledge. The lagged parameters indicated that the age-associated influence of speed on knowledge was greater than the converse effect of knowledge on speed.

Klumb (2001) used MLM to evaluate intraindividual time-sampling data. Older participants provided up to 30 self-reports of activities, perceived difficulty of activities, affect, and self-efficacy ratings over a 6-day period. MLM was then used to estimate relations among level and change in different variables. Self-efficacy level predicted intraindividual variability in positive affect and perceived task difficulty.

From Statics to Dynamics: Recent Developments

The history of scientific disciplines supports the argument that phenomena are not fully understood until the pertinent lawful relationships are written in dynamic, change-process-oriented terms (West, 1985). Behavioral and social science has not reached this level of development despite advocacy for dynamic approaches (e.g., Coleman, 1968). With rare exceptions, social and behavioral science continues to feature the static-equilibrium, stability-oriented conceptions held by most sciences until the past century (Holling, 1973). Thus, exciting developments in analysis procedures that attempt to represent change in newer, more powerful ways but that are not particularly dynamic—such as LGM—can be regarded as "way stations" in the evolution of change measurement (Molenaar et al., 2003).

Fortunately, we are seeing renewed interest in representing dynamic relationships, including aspects of change measurement, in powerful mathematical forms such as difference and differential equations (Arminger, 1986; Boker, 2001; McArdle & Hamagami, 2001; Tuma & Hannan, 1984). A variety of dynamical modeling thrusts have found their voices in the psychological literature of the past decade (e.g., Vallacher & Nowak, 1994; Thelen & Smith, 1994). Boker and his colleagues have demonstrated the versatility of the damped linear oscillator model with applications in the area of substance abuse among adolescents (e.g., Boker & Graham, 1998). Factor analysis has received a dynamical "face lift" within the context of linear dynamical systems modeling. Dynamic factor analysis (DFA) models—extensions of Cattell's P-technique factor analysis model—have been put forward (e.g., Molenaar, 1994; see Nesselroade & Ghisletta, 2003, and Nesselroade, McArdle, Aggen, & Meyers, 2002, for recent reviews).

Applications of these techniques are beginning to appear in the empirical literature. For example, Shifrin, Hooker, Wood, and Nesselroade (1997) used DFA to analyze daily self-reported affect scores for 70+ days by a sample of participants with Parkinson's disease. The analyses, which were done on each individual's data, revealed that patterns of intraindividual variability in reported affect varied across individuals both in terms of dimensionality (number of factors) and in terms of temporal organization (prediction of later affective status from earlier affective status). Thus, the notion of structure was elaborated from one of mere static dimensionality (number of factors) to include the idea of temporal organization (lagged relationships between factors and manifest variables). This latter conveys the dynamical relationships concerning how changes in the factors are linked to changes in the variables.

The fuller integration of dynamical modeling approaches into the study of behavioral change generally will be a difficult and prolonged task. In addition to the sheer inertia that must be overcome from decades of focusing on simple measures of change, such as difference scores, the use of dynamical modeling procedures is relatively data intensive. There are also difficult problems of estimation that must be resolved (e.g., Boker & Nesselroade,

2002) and the available computer software for conducting these analyses is not as generally available, nor is it as user friendly as some of the better conventional statistical packages. Despite the uncommon level of quantitative sophistication required to use these models (e.g., estimating first and second derivatives before the actual model fitting can take place) and to interpret the results, their promise is great. Indeed, the shift toward more dynamical representations relying on advanced mathematical forms characterizes how other scientific disciplines have evolved (West, 1985), and this trend is likely to influence developmental psychology as well.

Open Questions and Future Directions

The foregoing review of new statistical approaches to the measurement of change raises some critical issues that are inherently part of any longitudinal design and analysis targeting developmental change. Developmental psychologists—particularly ones interested in changes over the life span (e.g., Baltes, 1987, 1993; Schaie, 1996) have considered in some detail the problems associated with inferring age-related changes from empirical data (Nesselroade & Labouvie, 1985; Schaie, 1977). Inferences about aging require attention to a number of issues in design (e.g., Alwin & Campbell, 2001; Cook & Campbell, 1979; Schaie, 1977) and measurement (e.g., Baltes & Nesselroade, 1970; Labouvie, 1980; Schaie & Hertzog, 1985). However, these considerations have not always influenced applications of techniques like LGM or MLM to longitudinal data. To bring this perspective into clearer focus, in the subsections below, we summarize and discuss a few key issues about analyzing developmental change.

What Are the Appropriate Assumptions for Models of Change in Adulthood?

The quickest and most pragmatic longitudinal design for adult developmental research involves the use of longitudinal sequences (Baltes et al., 1988; Schaie, 1977), in which an initial cross-sectional sample is followed longitudinally and new cross-sectional samples are also collected (and followed longitudinally) at later points in time. Once a second occasion of measurement is available, models can be used to estimate change parameters (Ghisletta & Lindenberger, 2003), even though more cross-sectional than longitudinal information is available concerning age-related change. LGM and MLM models for the data, estimating an age-basis for the function, rely on the assumption of convergence (McArdle & Bell, 2000), that is, that age differences in cross-section and longitudinal changes can be converged to estimate age-related mean trends and individual differences in change around those trends.

Ironically, most of the literature in life-span developmental methodology challenges the viability of convergence assumptions not merely because of the well-known problem of cohort effects and/or Age $\times$ Cohort interactions (Kaufman, 2001; Neisser, 1998; Schaie, 1996) but also because of the implicit additional assumptions: missing-at-random data for those who are lost to longitudinal follow-up, absence of period, practice, and instrumentation confounds, and so on. Unfortunately, most applications of LGM and MLM to model longitudinal means and covariance structures do not evaluate the viability of convergence assumptions, nor do they explicitly attempt to estimate cohort effects from the data (but see Miyazaki & Raudenbush, 2000). In a minimal longitudinal sequence with a wide cross-sectional age span and a short retest interval, it may not be possible to empirically identify cohort differences and age changes, and different assumptions can lead to widely disparate outcomes (Donaldson & Horn, 1992). Nevertheless, users of advanced longitudinal models assume convergence at the peril of producing invalid estimates of age-related change.

A related assumption in longitudinal analysis is the assumption of ergodicity (Molenaar et al., 2003), that is, that intraindividual change functions can be captured by aggregating data in panel studies and studying individual differences in intraindividual change (as in the standard LGM and MLM analyses we have reviewed). Molenaar et al. (2003) argued that in many cases standard LGM models fail to capture important patterns of intraindividual variability because of aggregation bias. They challenged whether LGM parameters, such as the variance in Shapes, provides a valid description of individual differences in rates of change.

The nature of this bias may be influenced by simplistic assumptions about the form of the underlying developmental function. Some statistical models leave the developmental function open and relatively unconstrained (as in the polynomial growth curve approach), whereas others specify a particular functional form. Likewise, models vary with respect to the assumptions they carry regarding how the aggregate developmental function should organize measurement of change. The latent change model (McArdle & Nesselroade, 1994) assumes that individual differences in change can be analyzed when the mean change function is ignored, but the specification of the model implies that change is linear for the time interval spanned by the latent changes. Likewise, many MLM models implicitly model simplified developmental functions because they specify only a linear age term or linear and quadratic age terms.

In general, applications of models for change should do more to evaluate the assumptions that are (often implicitly) built into the models. In some cases, explicit evaluation of convergence properties can be encouraging. Hultsch et al. (1998) found that cross-sectional comparisons of cognitive factor structures for their young-old and old-old subsamples revealed age differences in factor correlations that were later mirrored in parameters from a latent change model. For example, correlations of other cognitive factors with Working Memory were higher in the old-old cross-sectional sample, and the longitudinal model showed strong correlations of changes in Working Memory with changes in cognitive variables. This outcome suggested that the cross-sectional correlational differences were indirectly reflecting longitudinal changes in relationships between variables.

Are There Qualitative Differences in Patterns of Change?

Adult development and aging creates unique problems in defining the reference population of interest. Life-span developmental theorists have long recognized that the causes of change in adulthood are not necessarily ontogenetic in nature. A variety of nonnormative influences can create qualitatively different subgroups, in terms of patterns of change (e.g., Baltes, 1993, Busse, 1969). The problem of external validity or generalizability for studies of aging does not refer to the general population at any point in time (e.g., the inception point of a longitudinal panel study) but rather to a population of persons aging from birth to death, where aging is an intraindividual phenomenon and the time and causes of death are random variables. From this point of view, there is no single reference population, but rather, implicit subpopulations developing in heterogeneous ways because of normative and nonnormative influences on developmental change. If population mortality is nonrandom with respect to psychological processes (e.g., if the hostility facet of personality and other behaviors are risk factors for coronary artery disease; Haney et al., 1996), then attrition is not merely a problem of positive bias due to missing information but also a problem of shifting population composition. Indeed, data from the Seattle Longitudinal Study show that prior longitudinal change in intellectual functioning is a risk factor for subsequent mortality (Bosworth et al., 1999).

Then the question becomes, To what reference population does the aggregate developmental function generalize, if any (Nesselroade & Labouvie, 1985)? Referring back to Figure 1, it is not difficult to imagine an aggregate curve that makes accurate probabilistic statements (e.g., what is the best guess for the performance level of a 50-year old?) but that is an accurate depiction of the intraindividual change function of few, if any, individuals. This issue is considered in some detail by Sliwinski, Hofer, and Hall (2003; Sliwinski, Hofer, Hall, et al., 2003), who show that the presence of subclinical Alzheimer's disease can alter MLM estimates of age-changes and individual differences in rates of change in a panel study. Further, they showed that organizing the basis function by time to diagnosis, rather than age, provided a better fit to the patient group and dramatically reduced estimated individual differences in rates of cognitive change.

The problem cannot always be addressed by identifying, even ex post facto, subgroups with diagnosis of nonnormative pathology. Typical LGM and MLM applications assume homogeneity of developmental function; that is, that the same developmental function applies to all members of the population, with individuals varying only in the values of the parameters specified (e.g., magnitude of Level and Shape parameters). Heterogeneity in developmental functions is allowed only within the constraints defined by the parametric equations, including variation according to predictors of change. Granted, a considerable degree of heterogeneity in developmental change is afforded by variance in a nonlinear Shape parameter, but it may still not be sufficient to adequately describe and explain individual differences in change over time.

In contrast, person-centered approaches to the study of development and change (Bergman, Magnusson, & El Khouri, 2003) attempt to identify empirically subgroups of individuals who share a common developmental pathway. By studying the differences in these developmental pathways and identifying precursors, covariates, and predictors, one hopes to build explanatory systems for the differences that are observed in developmental outcomes.

Aldwin et al. (2001) reported an excellent example of the use of cluster analysis to evaluate qualitatitive differences in change. They used MLM to estimate individual developmental trajectories for physical health and psychological symptoms for panel members of the Normative Aging Study. They then used standard cluster analysis techniques to identify subgroups differing in patterns of symptom change. The modal pattern of symptoms involved men with low, stable physical and psychological symptom levels. However, smaller subgroups were identified with distinct patterns of personality–symptom relationships. For example, older men with high levels of physical health problems that increased over time were higher in hostility and anxiety, were overweight, and smoked. Similar work by Liang et al. (2003), using MLM followed by cluster analysis, yielded similar results. They suggest that, in old age, there are different patterns of change in subjective health. They identified three types of change patterns, each of which was associated with a different trajectory of change in functional competence in everyday life.

Techniques are available that combine longitudinal change models with a means of identifying subtypes of persons with respect to change in simultaneous estimation approaches. Woodbury and Manton (1991) developed a technique, grade-of-membership analysis, that allows individuals to be characterized as members of fuzzy sets, according to a weight matrix that reflects degree of membership in an idealized type (e.g., Manton, Cornelius, & Woodbury, 1995). The technique can be used to classify persons with respect to heterogeneous patterns of change over time. Manton, Siegler, and Woodbury (1986) used the techniques to identify persons varying in patterns of intellectual development in adulthood. A subtype of individuals had excellent physical and mental health and maintained intellectual functioning over time, whereas others were characterized by decline and a high probability of later dementia diagnosis.

Muthen (2001) has developed an approach that combines qualitatitive types with LGM. It explicitly allows for qualitatitive differences in shapes of developmental functions and can estimate whether an individual belongs to a particular latent class of persons with different change patterns. A potential limitation of this approach is that nonnormal distributions can produce spurious evidence for embedded subtypes (Bauer & Curran, in press).

It is too early to tell whether the promise of typological approaches will lead to new empirical advances in our understanding of adult psychological development. A critical issue that such approaches must address is that they can recover differences in overall level of functioning that are just as easily (if not more easily) understood as quantitative differences in the Level and Shape parameters from latent growth curve analysis. An exciting area for future work will be to refine our understanding of when a typologically oriented analysis produces strong evidence of true qualitative differences in patterns of developmental change, as opposed to mere quantitative differences in an underlying function. This is a classic and enduring problem in developmental analysis (Baltes & Nesselroade, 1970; Wohlwill, 1973).

How Are Patterns of Developmental Change Reflected in Model Parameters?

We believe there is a compelling need for simulation studies that work up from patterns of developmental change toward understanding how such effects would be manifested in parameter estimates from the types of statistical models reviewed in this article. For example, we have little knowledge about how restriction of range effects toward the bottom of a measurement scale might influence estimated variance in latent slopes in an LGM or MLM analysis. At present it is not known what the potential costs and biases are for the kind of two-stage application of random-effects models described earlier. To what extent is the estimation of individual slopes (akin to factor score estimation), followed by other exploratory techniques, likely to produce estimates and inferences about psychological change that are similar to estimates and inferences obtained by a simultaneous multivariate approach that estimates covariances between slopes and prediction equations for slopes without two-stage estimation? There is also a major need for additional simulation work that tests the consequences of violating assumptions such as convergence or ergodicity for LGM and MLM applications. At present we know far too little about the robustness of parameter estimates, or about the robustness of substantive inferences about development drawn from longitudinal models. Finally, the potential controversy over whether tests of factorial invariance requires estimation of factor means would be best resolved by analyzing simulation data where the possible consequences of selection on factor means and specific unique component means for invariance could be evaluated.

Reconceptualizing Measurement "Error"

Meredith and Horn (2001) argued that longitudinal analyses should explicitly model change in unique or specific factors, including the mean changes in these components. Their argument highlights the critical importance of conceiving of measurement error as consisting of multiple components that may or may not change in ways that are similar to the latent construct of interest. Measurement error in this sense is not random error, but rather a complex amalgam of multiple, systematic sources of variance that are treated as error because they are not necessarily the target of investigation. For example, individuals may choose a number on a rating scale for a variety of reasons that are independent of the target construct being rated, including social desirability, self-handicapping, aversion to extreme responses, and uncertainty about translations of scale. All such sources of error are lumped into the residual error term without an attempt to measure them explicitly. By assumption, and usually without supporting evidence, residual effects are treated as stochastic and isolated from the variables that are included in the model (the self-containment assumption; see James, Mulaik, & Brett, 1982).

In principle, these sources of error could behave differently over time than the constructs themselves, and the error components can have different impact on means and on variances. Models of processing component scores from experimental psychology are instructive in this regard. They indicate that not all components will have measurable means or measurable variances (Donaldson, 1983). In some cases experimental task manipulations have a substantial mean effect, but produce little variance (individual differences) in the magnitude of that effect. Latent difference score models for such experiments will not produce robust variance estimates for an experimental effect parameter (Donaldson, 1983; Embretson, 1983). By analogy, there could also be little variation in developmental change in such experimental effects. Some experimental task parameters (e.g., mental rotation task slopes) demonstrate reliable individual differences, but others do not. So the measurement error problem must be evaluated in the context of each specific task, or in the case of development, in the context of specific constructs and situations across parts of the life span.

The implication is that scientists taking to heart Meredith and Horn's (2001) recommendation to model means and variances of changes in specific components must allow for the possibility that the specific error components that predict individual differences in change may not produce average mean changes, and vice versa. Thus, components that explain changes in observed variable means may not account for individual differences in observed changes over time. The typical LGM specification presumes, to the contrary, that mean changes and changes in covariance structures are determined by the same component processes. Generally, standard models for longitudinal data incorporate assumptions about measurement error that seem simplistic and even inaccurate. Univariate LGMs typically assume that error variance at each longitudinal measurement occasion is homogeneous and orthogonal to the growth process being modeled. Yet error, in this sense, is merely residual variance not accounted for by the growth curve, and as such can contain multiple systematic sources of variance. Specifying homogeneity of error variance may be required to identify the model and/or to obtain stable parameter estimates, but this does not imply that the assumption is correct or that the estimates obtained thereby are unbiased.

Summary and Conclusions

The methodological arsenal of scientists interested in aging is undergoing constant revision, and major progress is being made in providing new tools and approaches for statistical models of developmental change in adulthood. One of our concerns, at present, is that (to borrow a phrase from our late colleague Jack Wohlwill), we cannot allow the statistical wing to flap the scientific bird. Choice of modeling technique and implementation of a specific modeling approach should be grounded in and reflect both the theoretical nature of the developmental phenomenon and the features of the sampling design that selected persons, variables, and contexts for empirical observation. We advocate renewed attention to the issues of internal validity, construct validity, and statistical conclusion validity (Cook & Campbell, 1979) in research on developmental change. Sophistication in modeling application is not merely a function of the elegance of the statistical model. It is also manifested in how one matches models to phenomena. Optimally, one does so in a way that aggregates at minimal cost, that represents key concepts as model parameters capturing the dynamics of change, and that allows for the complexity of multiple patterns and multiple influences on change.

References

Alder, A. G., Adam, J., & Arenberg, D. (1990). Individual-differences assessment of the relationship between change in and initial level of adult cognitive functioning. *Psychology and Aging, 5,* 560–568.

Aldwin, C. M., Spiro, A., III, Levenson, M. R., & Bosse, R. (1989). Longitudinal findings from the Normative Aging Study: I. Does mental health change with age? *Psychology and Aging, 4,* 295–306.

Aldwin, C. M., Spiro, A., III, Levenson, M. R., & Cupertino, A. P. (2001). Longitudinal findings from the Normative Aging Study: III. Personality, individual health trajectories, and mortality. *Psychology and Aging, 16,* 450–465.

Alwin, D. F. (1988). Structural equation models in research on human development and aging. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 71–169). New York: Springer Publishing Company.

Alwin, D. F., & Campbell, R. T. (2001). Quantitative approaches: Longitudinal methods in the study of human development and aging. In R. H. Binstock & L. K. George (Eds.), *Handbook of aging and the social sciences* (pp. 22–43). San Diego, CA: Academic Press.

Alwin, D. F., & Krosnick, J. A. (1989). Aging and susceptibility to attitude change. *Journal of Personality and Social Psychology, 57,* 416–425.

Anstey, K. J., Hofer, S. M., & Luszcz, M. A. (2003). A latent growth curve analysis of late-life cognitive and sensory function over 8 years: Evidence for specific and common factors underlying change. *Psychology and Aging, 18,* 714–726.

Arminger, G. (1986). Linear stochastic differential equation models for panel data with unobserved variables. In N. Tuma (Ed.), *Sociological methodology 1986* (pp. 187–212). San Francisco: Jossey-Bass.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology, 23,* 611–626.

Baltes, P. B. (1993). The aging mind: Potential and limits. *Gerontologist, 33,* 580–594.

Baltes, P. B., & Labouvie, G. V. (1973). Adult development of intellectual performance: Description, explanation, and modification. In C. Eisdorfer & M. P. Lawton (Eds.), *The psychology of adult development and aging* (pp. 157–219). Washington, DC: American Psychological Association.

Baltes, P. B., & Nesselroade, J. R. (1970). Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. *Developmental Psychology, 2,* 163–168.

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.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1988). *Life-span developmental psychology: Introduction to research methods.* Hillsdale, NJ: Erlbaum.

Baltes, P. B., Staudinger, U., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology, 50,* 471–507.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural equation modeling. *Sociological Methods and Research, 16,* 78–117.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3–20). Madison: University of Wisconsin Press.

Berg, S. (1996). Aging, behavior, and terminal decline. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 323–337). San Diego, CA: Academic Press.

Bergman, L. R., Magnusson, D., & El Khouri, B. M. (2003). *Studying individual development in an interindividual context: A person-oriented approach.* Mahwah, NJ: Erlbaum.

Bloom, B. S. (1964). *Stability and change in human characteristics.* New York: Wiley.

Boker, S. M. (2001). Differential structural equation modeling of intraindividual variability. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 5–27). Washington, DC: American Psychological Association.

Boker, S. M., & Graham, J. (1998). A dynamical systems analysis of adolescence substance abuse. *Multivariate Behavioral Research, 33,* 479–507.

Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave data. *Multivariate Behavioral Research, 37,* 127–160.

Bosworth, H. B., Schaie, K. W., & Willis, S. L. (1999). Cognitive and sociodemographic risk factors for mortality in the Seattle Longitudinal Study. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 54,* 273–282.

Bradburn, N. M. (1969). *The structure of psychological well-being.* Chicago: Aldine.

Brickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intellect across the adult life span. *Intelligence, 20,* 309–328.

Browne, M. W., & duToit, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47–68). Washington, DC: American Psychological Association.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147–158.

Busse, E. W. (1969). Theories of aging. In E. W. Busse & E. Pfeiffer (Eds.), *Behavior and adaptation in late life* (pp. 11–32). Boston: Little, Brown.

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

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: Guilford Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cattell, R. B. (1966). Patterns of change: Measurement in relation to state dimension, trait change, lability, and process concepts. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 355–402). Chicago: Rand McNally.

Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychophysical source traits in a normal individual. *Psychometrika, 12,* 267–288.

Christensen, H., Hofer, S. M., Mackinnon, A. J., Korten, A. E., Jorm, A. F., & Henderson, A. S. (2001). Age is no kinder to the better educated: Absence of an association established using latent growth techniques in a community sample. *Psychological Medicine, 31,* 15–27.

Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., Jacomb, P., & Rodgers, B. (1999). An analysis of diversity in cognitive performance of elderly community dwellers: Individual differences in change scores as a function of age. *Psychology and Aging, 14,* 365–379.

Coleman, J. S. (1968). The mathematical study of change. In H. M. Blaylock, Jr., & A. Blaylock (Eds.), *Methodology in social research* (pp. 428–478). New York: McGraw-Hill.

Collins, L. M., & Horn, J. L. (1991). *Best methods for the analysis of change.* Washington, DC: American Psychological Association.

Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New methods for the analysis of change.* Washington, DC: American Psychological Association.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Costa, P. T., Jr., & McCrae, R. R. (1994). Set like plaster: Evidence for the stability of adult personality. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 21–40). Washington, DC: American Psychological Association.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 8–80.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107–135). Washington, DC: American Psychological Association.

Donaldson, G. (1983). Confirmatory factor analysis models of information processing stages: An alternative to difference scores. *Psychological Bulletin, 94,* 143–151.

Donaldson, G., & Horn, J. L. (1992). Age, cohort, and time developmental muddles: Easy in practice, hard in theory. *Experimental Aging Research, 18,* 213–222.

Duncan, T. E., & Duncan, S. E. (1995). Modeling the process of development via latent growth curve methodology. *Structural Equation Modeling, 2,* 187–213.

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences.* New York: Oxford University Press.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Estes, W. K. (1956). The problem of inferences from curves based on group data. *Psychological Bulletin, 53,* 134–140.

Finch, J. F., Jr., & Zautra, A. J. (1992). Testing latent longitudinal models of social ties and depression among the elderly: A comparison of distribution-free and maximum likelihood estimates with nonnormal data. *Psychology and Aging, 7,* 107–118.

Finkel, D., Reynolds, C. A., McArdle, J. J., Gatz, M., & Pedersen, N. L. (2003). Latent growth curve analyses of accelerated decline in cognitive abilities in late adulthood. *Developmental Psychology, 39,* 535–550.

Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging, 18,* 696–713.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). New York: Halsted Press.

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

Hamagami, F., & McArdle, J. J. (2001). Advanced studies of individual differences linear dynamic models for longitudinal data analysis. In G. A. Marcoulides & R. W. Schumacker (Eds.), *New developments and techniques in structural equation models* (pp. 203–246). Mahwah, NJ: Erlbaum.

Hampson, E. (1990). Variations in sex-related cognitive abilities across the menstrual cycle. *Brain and Cognition, 14,* 26–43.

Haney, T. L., Maynard, K. E., Houseworth, S. J., Scherwitz, L. W., Williams, R. B., & Barefoot, J. C. (1996). Interpersonal hostility assessment technique: Description and validation against the criterion of coronary artery disease. *Journal of Personality Assessment, 66,* 386–401.

Harris, C. W. (Ed.). (1963). *Problems in measuring change.* Madison: University of Wisconsin Press.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.

Hertzog, C., & Bleckley, M. K. (2001). Age differences in the structure of intelligence: Influences of information processing speed. *Intelligence, 29,* 191–217.

Hertzog, C., Cooper, B. P., & Fisk, A. D. (1996). Aging and individual differences in the development of skilled memory search performance. *Psychology and Aging, 11,* 497–520.

Hertzog, C., & Dixon, R. A. (1996). Methodological issues in research on cognition and aging. In F. Blanchard-Fields & T. Hess (Eds.), *Perspectives on cognitive change in adult development and aging* (pp. 66–121). New York: McGraw-Hill.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain and Language, 42,* 248–269.

Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging, 18,* 755–769.

Hertzog, C., Hultsch, D. F., & Dixon, R. A. (1989). Evidence for the convergent validity of two self-report metamemory questionnaires. *Developmental Psychology, 25,* 687–700.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait–state distinction for the structural modeling of developmental change. *Child Development, 58,* 93–109.

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: I. Analysis of longitudinal covariance structures. *Psychology and Aging, 1,* 159–171.

Hertzog, C., Van Alstine, J., Usala, P. D., Hultsch, D. F., & Dixon, R. A. (1990). Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment, 2,* 64–72.

Hofer, S. M., Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., & Easteal, S. (2002). Changes in cognitive functioning associated with apoE genotype in a community sample of older adults. *Psychology and Aging, 17,* 194–208.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics, 4,* 1–23.

Horn, J. L., & Donaldson, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist, 31,* 701–709.

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

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged.* New York: Cambridge University Press.

Hultsch, D. F., Hertzog, C., Small, B. J., & Dixon, R. A. (1999). Use it or lose it? Engaged life style as a buffer of cognitive decline in aging. *Psychology and Aging, 14,* 245–263.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 57,* P101–P115.

Hummel-Rossi, B., & Weinberg, S. L. (1975). Practical guidelines in applying current theories to the measurement of change. Part I: Problems in measuring change and recommended procedures. Part II: Numerical illustrations of research designs concerned with measuring change [Abstract]. *Journal Abstract Supplement Service Catalog of Selected Documents in Psychology, 5,* 226.

James, L. R., Mulaik, S. A., & Brett, J. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–352). New York: Academic Press.

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity, 24,* 387–404.

Kaufman, A. S. (2001). WAIS–III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence, 29,* 131–167.

Kemper, S., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging, 16,* 600–614.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kessler, R. C., & Greenberg, D. F. (1981). *Linear panel analysis: Models of quantitative change.* New York: Academic Press.

Klumb, P. (2001). Action regulation in old age: Time samples from everyday contexts. *International Journal of Behavioral Development, 25,* 392–399.

Labouvie, E. W. (1980). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s: Selected contemporary issues in the psychology of aging* (pp. 493–502). Washington, DC: American Psychological Association.

Lamiell, J. T. (1997). Individuals and the differences between them. In R. Hogan, J. A. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 117–141). New York: Academic Press.

Lane, C. J., & Zelinski, E. M. (2003). Longitudinal hierarchical linear models of the Memory Functioning Questionnaire. *Psychology and Aging, 18,* 38–53.

Liang, J. (1985). A structural integration of the Affect Balance Scale and the Life Satisfaction Index A. *Journal of Gerontology, 40,* 552–561.

Liang, J., Shaw, B. A., Krause, N. M., Bennett, J. M., Blaum, C., Kobayashi, E., et al. (2003). Changes in functional status among older adults in Japan: Successful and usual aging. *Psychology and Aging, 18,* 684–695.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192–211.

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

Lynch, S. M., & George, L. K. (2002). Interlocking trajectories of loss-related events and depressive symptoms among elders. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 57,* P117–P125.

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

MacDonald, S. W. S., Hultsch, D. F., Strauss, E., & Dixon, R. A. (2003). Age-related slowing of digit symbol substitution revisted: What do longitudinal age changes reflect? *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 58,* P187-P194.

Maitland, S. B., Dixon, R. A., Hultsch, D. F., & Hertzog, C. (2001). Well-being as a moving target: Measurement equivalence of the Bradburn Affect Balance Scale. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 56,* P69–P77.

Manton, K. G., Cornelius, E. S., & Woodbury, M. A. (1995). Nursing home residents: A multivariate analysis of their medical, behavioral, psychosocial, and service use characteristics. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences, 50,* M242–M251.

Manton, K. G., Siegler, I. C., & Woodbury, M. A. (1986). Patterns of intellectual development in late life. *Journal of Gerontology, 41,* 486–499.

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

McArdle, J. J., & Aber, M. (1990). Patterns of change within latent variable structural equation models. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Vol. 1. Principles and structuring change* (pp. 151–224). San Diego, CA: 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* (Vol. 2, pp. 21–44). San Diego, CA: Academic Press.

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

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

McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal data using latent growth structural equation models. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 276–304). Washington, DC: American Psychological Association.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–175). Washington, DC: American Psychological Association.

McArdle, J. J., Hamagami, F., Elias, M. F., & Robbins, M. A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research, 17,* 29–51.

McArdle, J. J., & Nesselroade, J. R. (1994). Structuring data to study development and change. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological innovations* (pp. 223–268). Hillsdale, NJ: Erlbaum.

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

McAuley, E., Katula, J., Mihalko, S. L., Blissmer, B., Duncan, T. E., Pena, M., & Dunn, E. (1999). Mode of physical activity and self-efficacy in older adults: A latent growth curve analysis. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 54,* P283-P292.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5,* 23–43.

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

Meredith, W., & Horn, J. L. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 201–240). Washington, DC: American Psychological Association.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods, 5,* 44–63.

Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 155–180). Thousand Oaks, CA: Sage.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 339–360). Dordrecht, the Netherlands: Kluwer Academic.

Muthen, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: American Psychological Association.

Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Palermo (Eds.), *Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.

Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. Heatherton & J. Weinberger (Eds.), *Can personality change*? Washington, DC: American Psychological Association.

Nesselroade, J. R., & Ghisletta, P. (2003). Structuring and measuring change over the lifespan. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 317–337). Dordrecht, the Netherlands: Kluwer Academic.

Nesselroade, J. R., & Labouvie, E. W. (1985). Experimental design in research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 35–60). New York: Van Nostrand Reinhold.

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Alternative dynamic factor models for multivariate time-series analyses. In D. M. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Advances and techniques* (pp. 235–265). Mahwah, NJ: Erlbaum.

Nesselroade, J. R., & McCollam, K. M. S. (2000). Putting the process in developmental process. *International Journal of Behavioral Development, 24,* 295–300.

Nesselroade, J. R., & Molenaar, P. C. M. (2003). Quantitative models for developmental processes. In J. Valsiner & K. Connolly (Eds.), *Handbook of developmental psychology* (pp. 622–639). London: Sage.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88,* 622–637.

Newsom, J. T., Nishishiba, M., Morgan, D. L., & Rook, K. S. (2003). The relative importance of three domains of positive and negative social exchanges: A longitudinal model with comparable measures. *Psychology and Aging, 18,* 746–754.

Rabbitt, P., Diggle, P., Smith, D., Holland, F., & McInnes, L. (2001). Identifying and separating the effects of practice and of cognitive aging during a large longitudinal study of elderly community residents. *Neuropsychologia, 39,* 532–543.

Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 35–64). Washington, DC: American Psychological Association.

Raykov, T. (1999). Are simple change scores obsolete? An approach to studying correlates and predictors of change. *Applied Psychological Measurement, 23,* 120–126.

Reinert, G. (1970). Comparative factor analytic studies of intelligence through the human life-span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 468–485). New York: Academic Press.

Reynolds, C. A., Finkel, D., Gatz, M., & Pedersen, N. L. (2002). Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth curve models. *Experimental Aging Research, 28,* 407–433.

Reynolds, C. A., Gatz, M., & Pedersen, N. L. (2002). Individual variation in cognitive decline: Quantitative methods for describing patterns of change. *Psychology and Aging, 17,* 271–287.

Riegel, K. F., & Riegel, R. M. (1972). Development, drop, and death. *Developmental Psychology, 6,* 309–316.

Rogosa, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press.

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88,* 245–258.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 114,* 726–748.

Rovine, M. J., & Molenaar, P. C. M. (1999). The covariance between level and shape in the latent growth curve model: An explanation based on a simplified LISREL model. *Methods of Psychological Research, 3,* 95–107.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Rovine, M. J., & Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 67–93). Washington, DC: American Psychological Association.

Rudinger, G., & Reitz, C. (2001). Structural equation modeling in longitudinal research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 29–52). San Diego, CA: Academic Press.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Salthouse, T. A. (2000). Methodological assumptions in cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 467–498). Mahwah, NJ: Erlbaum.

SAS Institute, Inc. (2000). *SAS/STAT user's guide, Version 8.* Cary, NC: Author.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 177–200). Washington, DC: American Psychological Association.

Schaie, K. W. (1977). Quasi-experimental designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 39–58). New York: Van Nostrand Reinhold.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.

Schaie, K. W., & Hertzog, C. (1985). Measurement in the psychology of adulthood and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 61–92). New York: Van Nostrand Reinhold.

Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in aging research. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 53–77). San Diego, CA: Academic Press.

Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology and Aging, 13,* 8–20.

Schaie, K. W., Willis, S. L., Hertzog, C., & Schulenberg, J. E. (1987). Effects of cognitive training on primary mental ability structure. *Psychology and Aging, 2,* 233–242.

Schooler, C., Mulatu, M. S., & Oates, G. (1999). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging, 14,* 483–506.

Shifren, K., Hooker, K. A., Wood, P. K., & Nesselroade, J. R. (1997). Structure and variation of mood in individuals with Parkinson's disease: A dynamic factor analysis. *Psychology and Aging, 12,* 328–339.

Sliwinski, M., & Buschke, H. (1999). Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychology and Aging, 14,* 18–33.

Sliwinski, M. J., Hofer, S. M., & Hall, C. (2003). Correlated and coupled cognitive change in older adults with and without preclinical dementia. *Psychology and Aging, 18,* 672–683.

Sliwinski, M. J., Hofer, S. M., Hall, C., Buschke, H., & Lipton, R. B. (2003). Modeling memory decline in older adults: The importance of preclinical dementia, attrition, and chronological age. *Psychology and Aging, 18,* 658–671.

Small, B. J., Hertzog, C., Hultsch, D. F., & Dixon, R. A. (2003). Instability after only 6 years? A longitudinal confirmatory factor analysis of the NEO personality inventory. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 53,* P166–P176.

Spiro, A., III, Aldwin, C. M., Levenson, M. R., & Bosse, R. (1990). Longitudinal findings from the Normative Aging Study: II. Do emotionality and extraversion predict symptom change? *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 45,* P136–P144.

Stricker, L. J., & Rock, D. A. (1987). Factor structure of the GRE general test in young and middle adulthood. *Developmental Psychology, 23,* 233–242.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Tisak, J., & Meredith, W. (1990). Longitudinal factor analysis. In A. von

Eye (Ed.), *Statistical methods in longitudinal research* (Vol. 1, pp. 125–149). San Diego, CA: Academic Press.
Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* New York: Academic Press.
Vallacher, R. R., & Nowak, A. (1994). *Dynamical systems in social psychology.* San Diego, CA: Academic Press.
Weatherell, J. L., Reynolds, C. A., Gatz, M., & Pedersen, N. (2002). Anxiety, cognitive performance, and cognitive decline in normal aging. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 57,* P246-P255.
West, B. (1985). *An essay on the importance of being nonlinear.* Berlin: Springer-Verlag.
Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17,* 179–193.
Wohlwill, J. F. (1973). *The study of behavioral development.* New York: Academic Press.
Wohlwill, J. F. (1991). The partial isomorphism between developmental theory and methods. *Annals of Theoretical Psychology, 6,* 1–43.
Woodbury, M. A., & Manton, K. G. (1991). Empirical Bayes approaches to multivariate fuzzy partitions. *Multivariate Behavioral Research, 26,* 291–321.
Zautra, A. J., Guarnaccia, C. A., & Reich, J. W., Jr. (1988). Factor structure of mental health measures for older adults. *Journal of Consulting and Clinical Psychology, 56,* 514–519.
Zelinski, E. M., & Lewis, K. L. (2003). Adult age differences in multiple cognitive functions: Differentiation, dedifferentiation, or process-specific change? *Psychology and Aging, 18,* 727–745.
Zimprich, D. (2002). Cross-sectionally and longitudinally balanced effects of processing speed on intellectual abilities. *Experimental Aging Research, 28,* 231–251.

Received August 6, 2002
Revision received June 3, 2003
Accepted July 7, 2003 ■