Psychological Methods
1997, Vol. 2, No. 4, 403–435

1082-989X/97/$3.00

Expanding Test–Retest Designs to Include Developmental Time-Lag Components

John J. McArdle and Richard W. Woodcock
University of Virginia

Test–retest data can reflect systematic changes over varying intervals of time in a "time-lag" design. This article shows how latent growth models with planned incomplete data can be used to separate psychometric components of developmental interest, including internal consistency reliability, test-practice effects, factor stability, factor growth, and state fluctuation. Practical analyses are proposed using a structural equation model for longitudinal data on multiple groups with different test–retest intervals. This approach is illustrated using 2 sets of data collected from students measured on the Woodcock–Johnson—Revised Memory and Reading scales. The results show how alternative time-lag models can be fitted and interpreted with univariate, bivariate, and multivariate data. Benefits, limitations, and extensions of this structural time-lag approach are discussed.

Test–retest data are often collected to examine test reliability and trait stability. In the traditional test–retest design, participants are measured on a battery of tests and then, at some specific interval of time, the same participants are measured again on the same tests. Test–retest data are often collected over short periods of time to examine the test–retest reliability of a test or a battery of tests (e.g., Stanley, 1971). When data have been collected over longer intervals of time, the stability of the trait is highlighted and the terms *longitudinal* and *panel analyses* are used (e.g., see Nesselroade & Baltes, 1979).

Researchers interested in the reliability or stability of a psychological attribute often report a test–retest correlation for a specific test. This correlation can be informative under certain traditional assumptions about the test and the persons under study. But this correlation can be misleading when these persons change in a nonrandom or systematic way during the interval of time between test and retest. Tests measuring traits that change over time will demonstrate lowered test–retest correlations. These effects on the correlation can come as a result of the short-term impacts of practice and retention or from longer term impacts of growth or maturation. In these cases, the results from test–retest studies confound concepts of trait stability with test reliability, and the quality usefulness of the tests may be compromised. These issues are well known in psychometric theory (e.g., Anastasi, 1954; Cattell, 1957; Gulliksen, 1950; Nunnally, 1978; Traub, 1994), but few studies have overcome these fundamental test–retest problems.

A great deal of research has demonstrated how it is possible, even advantageous in some cases, to estimate some developmental within-person variation from complete longitudinal information (e.g., see Nesselroade & Baltes, 1979). These growth models work best when large numbers of participants are measured at many occasions on many variables, but this kind of data collection is often not possible (Cohen, 1991). Thus, various alternative models have been used to analyze incomplete longitudinal convergence or cohort-sequential data (Bell, 1953; Schaie,

The research was supported by Grant AG-07137 from the National Institute on Aging.

Portions of this paper were presented at the annual meetings of the Society of Multivariate Experimental Psychology, Albuquerque, New Mexico, October 1991; the American Psychological Association, San Francisco, August 1991; and the Psychometric Society, Berkeley, California, June 1993.

We thank many colleagues for their comments and suggestions on this manuscript, especially Steve Aggen, Steve Boker, Aki Hamagami, John Horn, Patricia Hulick, John Nesselroade, and Carol Prescott.

Correspondence concerning this article should be addressed to John J. McArdle or Richard W. Woodcock, Department of Psychology, University of Virginia, Charlottesville, Virginia 22903.

1965) and some of these have used new techniques in linear structural equation modeling (SEM) techniques (e.g., Aber & McArdle, 1991; Anderson, 1993; Horn & McArdle, 1980; McArdle & Anderson, 1990; McArdle & Hamagami, 1992). These structural models provide some consideration for data that differ in interval of time between repeated testings—here called the "time-lag." The main purpose of this article is to illustrate this time-lag methodology and show how it can be used to enhance the usefulness of the traditional test–retest design.

This article is organized into several sections. First, we present a brief introduction on components of change with test–retest data. Second, a time-lag design is introduced and illustrated with two samples of time-lag data: (a) a short-term univariate study of memory, and (b) a longer term multivariate study of memory and reading. Third, we introduce a latent growth model and show how it can be used to examine changes over time-lags. Fourth, we discuss model estimation to illustrate how and why these longitudinal models can be fitted using only two occasions of measurement on each person in a planned incomplete data design. Univariate structural results for both data sets are based on using standard SEM computer programs (e.g., LISREL, Mx). For clarity, only key technical issues are discussed and technical notes are presented in the Appendix. We then expand the time-lag model to include a multivariate organization of growth and change, and we present some multivariate results from the second data set. Finally, we discuss some benefits, limitations, and extensions of this time-lag methodology.

Internal Consistency and Stability

Basic psychometric concepts of reliability are often formulated from studies with "split-half" or "parallel forms" (see Gullikson, 1950; Nunnally, 1978; Stanley, 1971). In these classical test theory models, it is assumed that the observation of a participant's test performance (here termed Y) reflects an underlying construct. This construct may be termed either a trait or a true score or a common factor (termed F). We also assume that any test score includes some error of measurement (termed E), and these errors are typically assumed to be independent of the true scores. More recent work in item response theory modeling (IRT) also retains many of these traditional assumptions (e.g., Reise, Widaman, & Pugh, 1993; cf. Vinsonhaler & Meredith, 1966).

An illustration of a true-score model for two alternative forms is presented in the first path diagram of Figure 1A. In this path diagram, squares are used to represent observed or measured variables (Y_a and Y_b); circles are used to represent unobserved or theoretical variables (e.g., E_a, E_b, and F); one-headed arrows are used to represent regression coefficients (e.g., $F \leftarrow Y_a$ is fixed at a value of 1); and two-headed arrows are used to represent estimated correlation or variance terms (e.g., $F \leftrightarrow F = V_f$). More details on these kinds of path diagrams are given by McArdle and Boker (1990).

These path diagrams can be used to illustrate many classical test theory results (see Traub, 1994). For example, the variance of the observed score (V_y) may be calculated as the sum of the variance of the trait (V_f) plus the variance due to error (V_e). In this simple model the ratio of the variance due to the trait (V_f) compared with the total variance of the observed test performance (V_y) can be written as a standardized variance component or ratio ($V_f^* = V_f/V_y$). Under these assumptions, the expected variance component V_f^* is identical to the observed correlation between measures ($R_{a,b}$) for split-half, alternative form, or parallel form data. This coincides closely with one common definition of the internal-consistency reliability ratio (R_{ic}), and this definition is often used as a primary index of the quality and usefulness of a test.

Stability and Change

The same measure may be observed on more than one occasion. In Figure 1B the time of measurement is indexed within brackets ($Y[1]$ and $Y[2]$). In this test–retest model, we include the covariance over time ($C_{y[1,2]}$) between the two factor scores, and this allows us to consider additional sources of variance in tests and traits. For instance, we may be interested in the factor-stability—the degree to which the common factor scores F remain the same over time. If we assume that the observed variances are equal over time, then the observed correlation over time ($R_{y[1,2]}$) is a direct index of factor stability. In this case, the factor-stability ratio and the internal-consistency ratio can be calculated using similar formulae (i.e., $R_{a,b} = R_{y[1,2]} = V_f^*$). However, these simple model assumptions might not hold, so the size of the factor stability is not necessarily an index of test quality. In more general terms, factor stability is a characteristic of development and change (Burr & Nesselroade, 1990; McCall,

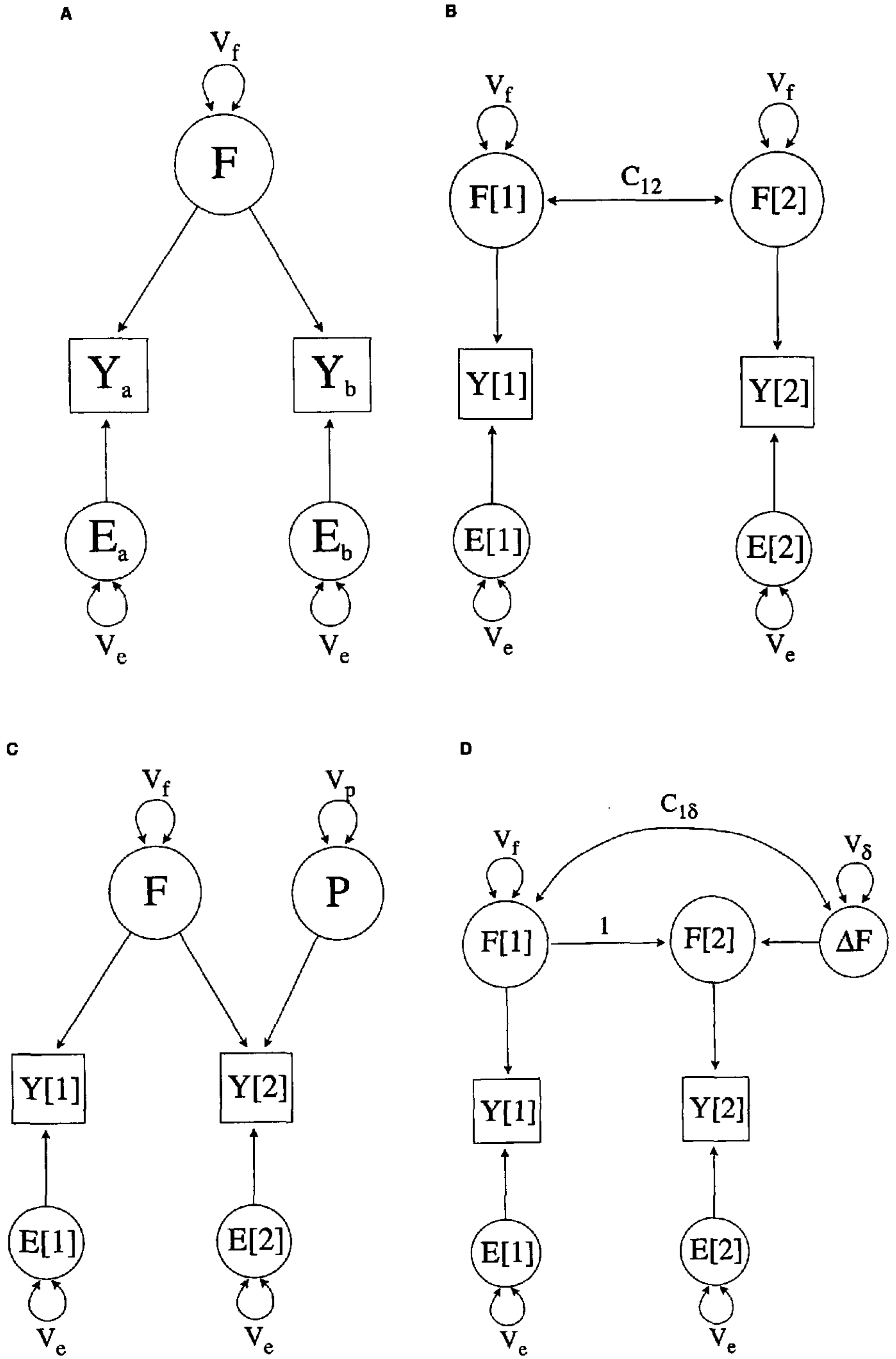

Figure 1. Alternative structural models for test–retest data. Square = observed variable; circle = unobserved variable; one-headed arrow = a unit-valued regression coefficient; two-headed arrow = a variance or covariance term. Panel A: Alternative forms. Panel B: Test–retest covariance. Panel C: Test–retest with practice. Panel D: Difference in true scores.

Appelbaum, & Hogarty, 1973; Nesselroade, 1983; Wohlwill, 1973).

There are a variety of practical cases where the factor stability can provide different information from the internal-consistency reliability. When occasions are spaced at shorter lengths of time we often expect a test-practice or test-retention effect (see Horn, 1972; Hundal & Horn, 1977; Jones, 1962; Nunnally, 1978, pp. 233–234; Woodworth, 1938, pp. 50–68). In some cases, a practice effect may be conceived of as an independent source of individual differences that only impacts observed scores after the first occasion of measurement.

The path diagram of Figure 1C shows a model that looks like Figure 1A with an additional latent variable *P*. This additional influence can increase (or decrease) the variance of the test at a second occasion (by the variance labeled V_p). If this is a practice effect, or virtually any additional source of variance, the observed test–retest correlation ($R_{y[1,2]}$) will not equal the standardized factor stability (V_f^*). To clarify this distinction, we need some practical way to separate the factor stability from the test practice-retention variance. We also want to account for any increases or decreases in the mean score as a result of practice or retention of the specific test material. These basic issues are raised again in analyses presented later.

More Complex Growth and Change Components

These theoretical conceptions can be expanded on in a number of ways. If we measure individuals over a long enough period of time we may find individual differences in the growth or maturation in a trait. Cattell (1957, pp. 343–344) used the terms *trait changes* and *function fluctuation* to indicate nonrandom and systematic changes in the factor scores over time. We use the term *factor growth* in a similar way here (after Horn, 1972; McArdle & Hamagami, 1992; Nesselroade & Baltes, 1979).

Figure 1D is an illustration of one kind of alternative model for test–retest studies. This model includes a factor score at a first occasion ($F[1]$) and a score on the same factor at the second occasion ($F[2]$). We have included a regression arrow with a unit value between the two factor scores, and this leads to a latent difference score (ΔF_n) that is possibly correlated with the initial factor score (i.e., in algebraic terms, $F[2]_n = F[1]_n + \Delta F_n$ so, by definition, $\Delta F_n = F[2]_n - F[1]_n$; e.g., McArdle & Nesselroade, 1994). This model leads to a set of expectations about the observed scores over time and can be used to organize and test hypotheses about change in the true-scores. Recent controversies about the "reliability of change scores" can also be formalized using this kind of model (e.g., Burr & Nesselroade, 1990; Rogosa & Willett, 1985; Willett, 1988). In more advanced models these developmental concepts can be related to systematic changes in the observed group means over time (as in Meredith & Tisak, 1990; McArdle & Hamagami, 1992, 1996).

Even more complex differences between reliability and stability are based on the need to account for psychological states. The most obvious state concepts in mental test performance include temporary fatigue, anxiety, and impulsivity, although other constructs may be involved as well (Cattell, 1957, pps. 349, 639–640, 683). The influences of psychological states on psychological performances are often defined as transitory or temporary features of behavior (e.g., Nesselroade & Bartsch, 1977; cf. Steyer, 1989; Steyer, Schwenkmezger, & Auer, 1990). These influences may vary in a systematic way across different variables within a specific time of measurement but also vary between times of measurement. In a multivariate model, we may consider the common state as a characteristic of a common factor that only occurs within a specific time point. Cattell (1957, 1964) used the term *state fluctuation* to deal with these components, and he suggested some unique ways to calculate state variance from empirical data. We explore these issues in more detail later.

Adding Time-Lags to Test–Retest Measurement

The developmental concepts raised above are difficult to examine, even when multiple-occasion longitudinal data are collected. To add to these problems, longitudinal data are among the most difficult data to collect. Unintended retesting effects can occur over short periods of time, and unintended attrition of participants can occur over longer periods of time (see Cohen, 1991; Nesselroade & Baltes, 1979). In practice, financial resources and other aspects of project planning may limit a longitudinal investigation to only two occasions.

In a single-occasion study, each variable (Y) is usually summarized by two statistics—one mean (M_y) and one variance (V_y)—although other aspects of the distribution of Y may be informative as well. The developmental information obtained from only one time point of measurement is cross-sectional so any model components can only reflect individual differences between persons. Similarly, each variable in a two-occasion test–retest study is usually summarized by five summary statistics: two means ($M_{y[1]}$ and $M_{y[2]}$), two variances ($V_{y[1]}$ and $V_{y[2]}$), and one correlation ($R_{y[1,2]}$). Unfortunately, these typical test–retest statistics do not contain enough information for a nonarbitrary separation of some developmental effects defined earlier. For example, an effect of growth-maturation (G) and an effect of practice-retention (P) both lead to increases in the means and variances at a second time point, so there is no easy way to separate such components (e.g., Jones, 1962). This kind of confounding also occurs with three or more time points as well (see Heise, 1969).

It is possible to improve the usefulness of two time points of measurement if we consider variation in the time-lag between tests. That is, previous research has shown how some developmental concepts can be examined if we consider issues related to time between measurements. Some time-lag models were illustrated by Thorndike (1931) in a study of repeated IQ measurements taken at different lengths of time delays. Thorndike fitted a relatively sophisticated regression model to these data—z-transformed correlations were the dependent variable and the time-lag between tests was the independent variable—so he could estimate the intercept of this regression model as the "instantaneous reliability" (the intercept) and the "change in reliability" (the slope) due to the time delay (also see Hartigan & Wigdor, 1989). In another more complex model, Vinsonhaler & Meredith (1966) developed an item response model that allowed for systematic change due to practice effects in repeated testings. Only a few researchers have studied time-lags in the context of a designed experiment (e.g., Schlesselman, 1973; Woodworth, 1938). The broad theme of using some features of time-lags in modeling analyses is adapted and extended in the rest of this paper.

Structural Equation Modeling of Test–Retest Data

The models to be used here reflect a mixture of traditional logic in reliability and change analysis. SEM techniques have been used for several decades to estimate reliability and stability from longitudinal data, especially using three or more points in time (e.g., Blok & Saris, 1983; Werts, Breland, Grandy & Rock, 1980; Heise, 1969; Hertzog & Nesselroade, 1987; Jöreskog & Sörbom, 1979; McDonald, 1980, 1985; Raffalovich & Bohrnstedt, 1987; Steyer, 1989). More recently, these SEMs have been used to examine growth and maturation in longitudinal data (e.g., Bergmann, 1993; McArdle, 1988; McArdle & Aber, 1990; McArdle & Epstein, 1987; McArdle & Nesselroade, 1994; Meredith, 1991; Meredith & Tisak, 1990). These new longitudinal structural analyses offer more flexible features for structured variation over time and models for mean changes using the same developmental functions, and allow the addition of psychometric measurement model parameters.

In a broader sense, the resulting longitudinal models are similar to classical variance components models of generalizability theory (e.g., Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991; Whitman, 1988). In several ways, these models also resemble a multilevel or hierarchical modeling approach to longitudinal data analysis (e.g., Bryk & Raudenbush, 1993; McArdle & Hamagami, 1996). The concepts of planned incomplete data appear in many other research designs, including analysis of variance (ANOVA) and multidimensional scaling (MDS; for review, see McArdle, 1994). Building on this prior work, we now further develop models that can use the time-lag between two measurements of the same tests.

Time-Lag Data

Planning Time-Lags in Test–Retest Measurement

The models examined here are based on repeated measurements collected over a variety of time-lags. The layout of Figure 2 presents one potentially useful plan for test–retest measurement.

This layout mimics the traditional test–retest data with one key exception—in this model the time between the test and retest is defined by time-lag t, and this time-lag is not the same for each group studied. Each column defines an independent group based on the variables we do collect (the observed squares) and the variables we do not collect (the unobserved circles). By convention, a zero time-lag ($t = 0$) is used to indicate the initial time of measurement. Included in a first group (column 1) are persons measured only at $t = 0$ and $t = 1$. For this group, we assume variables at time $t > 1$ are not measured again and we treat these occasions as latent variables. In the second group (column 2) we include persons measured only at $t = 0$ and $t = 2$. Here the time-lag of $t = 2$ indicates some predefined interval of time (e.g., 2 months), and latent variables are used to indicate no

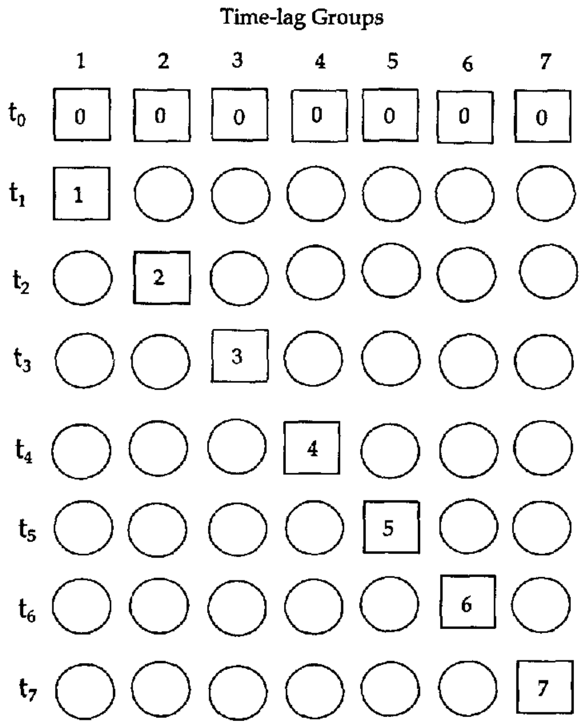

Figure 2. A time-lag (t) design for two-occasion test–retest data. Each column defines an independent group based on the variables collected (squares) from variables not collected (circles).

measurements at time $t = 1$ or at $t > 2$. The overall pattern of incomplete data shown here yields 7 independent groups, with the seventh group also measured twice, but at $t = 0$ and $t = 7$.

This planned time-lag layout of Figure 2 requires each participant to be measured at two occasions with a defined time-lag between tests. Where possible, we can accumulate the individuals into "time-lag groups" on the basis of a common unit(s) of time. The aggregation of persons into groups is not a formal necessity, and it requires several extra statistical assumptions (e.g., homogeneity of the persons, homogeneity of time-lag, etc.). This kind of aggregation will be used here mainly because it leads to convenient statistical displays and standard analyses.

We initially assume that there is no relationship between the scores at the first occasion ($Y[0]$) and the time-lag t chosen for each participant or group. This is a reasonable assumption when the time-lag between testings can be randomly assigned by the investigator and not selected by the participants. As it turns out, randomization to groups can be tested and may even be relaxed in more complex models. More flexible definitions of the time-lags can be based on substantive interest (e.g., 1 day, 1 month, etc.), equal intervals are not required, and more complex incomplete data collection plans can be effective (see Schlesselman, 1973; McArdle, 1994; McArdle & Hamagami, 1992, 1996).

Study 1: A Univariate Example of Daily Time-Lag Data

As a first illustration of a time-lag design we present data from a short-time test–retest study of the Woodcock–Johnson Psycho-Educational Battery—Revised (WJ–R; Woodcock & Johnson, 1989; McGrew, Werder, & Woodcock, 1991; see Appendix A). In this first study, 1,364 participants aged 5 to 19 were selected from the WJ–R standardization sample and participated in a short-term test–retest study of the WJ–R Memory-for-Names (MEMNAM) task. Scores on the first occasion were based on the number of "novel name–picture associations" held in memory at several points in the task. The retest occurred at a random selected time between 1 and 8 days later (with an average lag of about 3 days). At this second occasion these participants were asked to recall the picture associated with given names.

In theory, these data should show a pattern of change over time reflecting greater losses of memory with longer time delays. A univariate model will be developed in later sections to formally examine these ideas, and these models will be based on the time-lag data presented in Table 1 and Figure 3.

The top of Table 1 is a list of summary statistics for MEMNAM at two occasions. After adjusting these scores for the wide range of age differences (ages 5–19; see Appendix A) the means are centered at 0 (0.1 ± 12.6) at the first occasion, decline about 10 points (−9.9 ± 12.6) by the second occasion, and the overall (i.e., within age) test–retest correlation is $R_{y[1,2]} = .637$. These summary statistics are listed separately for the eight independent groups defined by the daily time-lag between tests. For each retest group, this list includes the sample size, the observed age-adjusted means, standard deviations, and test–retest correlations. Among many fluctuations, the means decrease, the deviations increase and decrease and the correlations are relatively high. The expected values listed in the bottom of Table 1 are the numerical results from a structural model and these will be discussed later.

In Figure 3 we present plots of the changes in these statistics over time for each variable. Figure 3A is a

Table 1
Sample Statistics for Eight Daily Time-Lag Groups Measured at Two Occasions on Woodcock–Johnson—Revised Memory for Names (N = *1,384*)

Days delay	N	Time 0 $M_{y[0]}$	Time 0 $D_{y[0]}$	Time t $M_{y[t]}$	Time t $D_{y[t]}$	Over time $(R_{y[0,t]})$
Observed time-lag scores (age residuals)						
1	279	−0.7	13.1	−8.7	13.9	.711
2	525	0.1	12.1	−8.8	11.7	.650
3	165	0.5	13.5	−8.8	13.1	.601
4	80	0.2	14.2	−12.6	13.4	.730
5	120	1.4	13.0	−10.8	11.6	.561
6	54	−3.4	12.3	−14.4	11.9	.546
7	137	1.4	10.9	−13.0	11.5	.536
8–14	24	0.6	9.7	−16.7	11.1	.546
Overall	1,384	0.1	12.6	−9.9	12.6	.637
Expected time-lag scores (from Model $\mathscr{C}_3$)						
1	279	0.1	12.6	−7.9	12.9	.676
2	525	0.1	12.6	−8.9	12.6	.662
3	165	0.1	12.6	−9.9	12.4	.645
4	80	0.1	12.6	−10.9	12.1	.625
5	120	0.1	12.6	−11.9	12.0	.601
6	54	0.1	12.6	−12.9	11.9	.574
7	137	0.1	12.6	−13.9	11.8	.545
8–14	24	0.1	12.6	−14.9	11.8	.513

plot of the initial means for each of the retest groups. Because of the random assignment, the initial $t = 0$ means (and deviations) should be statistically similar for all groups, but a smaller mean for Group 6 is apparent. Figure 3B is a plot of the mean difference between the two occasions within each time-lag group ($M_{y[t]} - M_{y[0]}$ for $t = 1$ to 8 groups). These mean differences are somewhat erratic but there are steady decreases over time. Figure 3C is a similar plot except here we depict the decreasing then increasing differences over time in the standard deviations ($D_{y[t]} - D_{y[0]}$). Finally, Figure 3D is a plot of the observed test–retest correlation over time ($R_{y[0,t]}$). These correlations decrease with an increasing time-lag except for the slightly higher correlation at Day 4 (for $N = 80$ with no outliers). These plots do not now include confidence boundaries for the wide range of sample sizes (ranging from $N = 525$ at $t = 2$ to only $N = 24$ at $t = 8$), but these will be estimated in later models.

Study 2: A Multivariate Example of Monthly Time-Lag Data

A different organization of time-lag data were collected in a longer term retest study of the WJ–R (see McGrew et al., 1991; the first section of the Appendix). In this second study, multivariate test–retest data were obtained on $N = 330$ persons aged 5–19 at time-lags between about 1 and 13 months. A test battery of 39 WJ–R tests was administered at the first occasion, and 26 of these tests were repeated at the second occasion. For illustrative analyses here we initially use four WJ–R raw scores: (MEMSEN or Y_1), Memory for Words (MEMWRD or Y_2), Letter–Word Identification (LWIDNT or Y_3), and Passage Comprehension (PSGCMP or Y_4). From these four variables we also created two standard WJ–R composite scores: a short-term memory factor (Memory or $[Y_1 + Y_2]/2$), and a broad reading cluster (Reading or $[Y_3 + Y_4]/2$).

A variety of substantive results are expected here. In theory, these two WJ–R composites (and four variables) indicate different developmental constructs that should have different patterns of change over time (see Horn, 1972, 1988; McGrew et al., 1991). The Memory scores represent the ability to hold new information in memory for short periods of time, and is thought to be largely independent of cultural knowledge and reasoning. In contrast, the Reading scores represent reading skills that are primarily learned from printed materials and training. On a substantive basis we expect that (a) the Memory scores may show some practice effects and minimal growth effects; (b) the Reading scores will show substantial growth effects but minimal practice effects; (c) similar growth patterns should exist within a pair representing the same composite; (d) different growth patterns should exist between variables of different composites; and (e) a single factor (g) model should not fit all these data very well. In the latter sections of this article, a multivariate model will be developed to deal with these ideas.

In Tables 2 and 3, we list the summary statistics for these two WJ–R composites at both occasions broken down into 11 separate groups defined by consecutive samples of $N = 30$. This aggregation leads to an unequal range of average time-lags ranging from 68 days to 424 days. Also listed are the age-adjusted means, standard deviations, and test–retest correlations for each retest group. The means and deviations generally increase over time, and the correlations are relatively high but show some fluctuations. The three columns of correlations listed alongside each variable and can be used to form the correlation matrix from both the Y and X variables at the specific time-lag. (Note that for each time-lag group in row t, the 3 correlations for Y in Table 2 can be combined with the 3 correlations for X in Table 3, and these 6 correla-

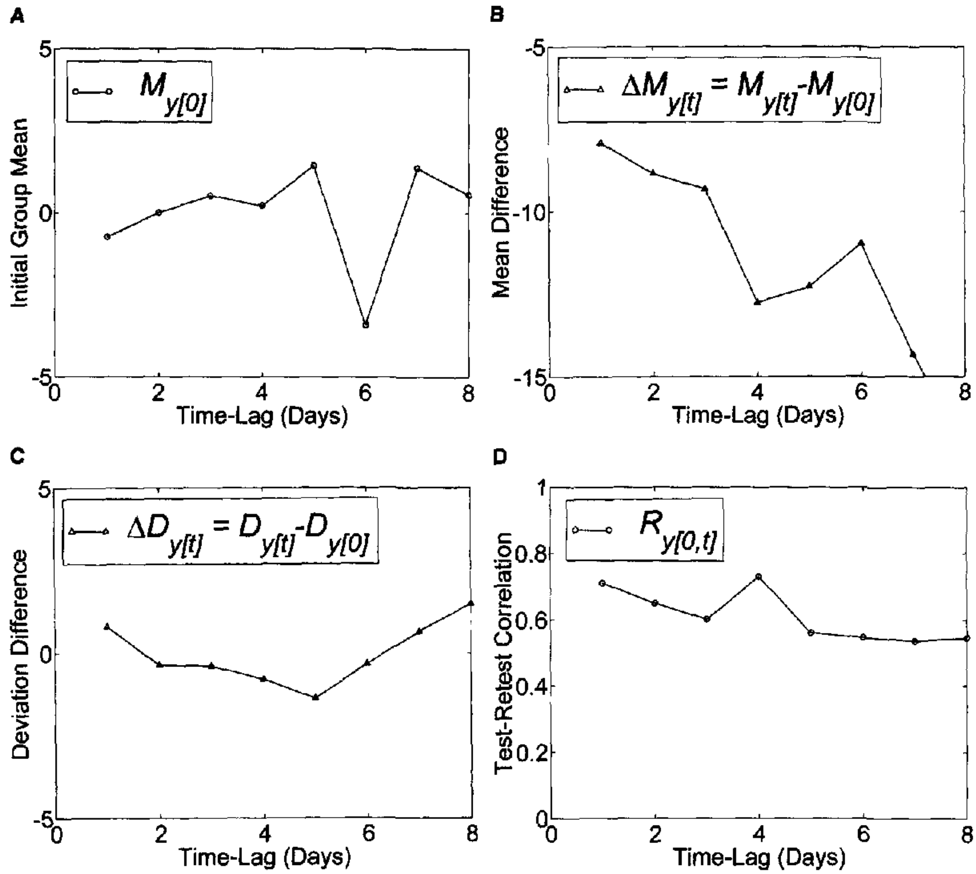

Figure 3. Observed statistics from Study 1: Woodcock–Johnson—Revised Memory for Names data from daily time-lag groups. Panel A: Means (M) at the initial time point. Panel B: Mean changes over time. Panel C: Deviation (D) changes over time. Panel D: Test–retest correlations (R).

tions are the unique elements of the 4 × 4 matrix of the variables $Y[0]$, $Y[t]$, $X[0]$, and $X[t]$ at the specific time-lag).

Figure 4 shows the changes in some statistics over time for each variable (with Memory as an "o" and Reading as an "x"). Figure 4A is a plot of the mean difference between the two occasions for each time-lag group ($\Delta M_{y[t]} - M_{y[0]}$ for $t = 1,11$). These means are erratic, but they do show small positive increases over time. The Memory means are initially higher and then decline over the year of time-lag. In Figure 4B we depict the differences over time in the standard deviations ($\Delta D_{y[t]} - D_{y[0]}$), and changes here are similar to the mean changes. Figure 4C is a plot of the test–retest correlations over time, and this erratic behavior may be an important source of misfit in later models. Figure 4D is a plot of the within-time correlation between these two variables over time, and these kinds of statistics will be used in later bivariate and multivariate time-lag analyses.

Time-Lag Models

A Time-Lag Structural Equation Model

Let us write a first structural model as

$$Y[t]_n = I_n + U[t]_n, \tag{1}$$

where $Y[t]$ is the observed variable score at some occasion t, the I is an unobserved initial level score, and the $U[t]$ is an unobserved unique or error score. In this model we assume I score is constant over time for each person and uncorrelated with the $U[t]$ score, and the $U[t]$ score is uncorrelated with other unique $U[t + l]$ scores at other times $t + l$ ($l > 0$). Under these assumptions we can expect a constant covariance (and correlation) over time for the observed scores $Y[t]$ and $Y[t + 1]$ at all intervals of time. This model is identical to the traditional true-score reliability model presented in Figure 1A, and this model can be useful when there is no systematic change over time.

Table 2
Sample Statistics for 11 Monthly Time-Lag Groups on Woodcock–Johnson—Short-Term Memory Composite Scores (Y = [MEMSEN + MEMWRD]/2)

Sample group	N	Time-lag (mean days)	Time 0 $M_{y[0]}$	Time 0 $D_{y[0]}$	Time t $M_{y[t]}$	Time t $D_{y[t]}$	Over time $R_{y[0,t]}$	Over time $R_{y[0],x[0]}$	Over time $R_{y[0],x[t]}$
1	30	68	−8.8	15.3	−4.9	17.6	.76	.51	.41
2	30	91	−5.6	13.0	3.5	13.3	.62	.29	.28
3	30	106	−8.9	16.1	−2.1	14.3	.78	.09	.24
4	30	139	4.0	18.2	5.9	15.3	.74	.16	.27
5	30	229	1.4	14.7	3.9	17.5	.87	.50	.65
6	30	268	−3.7	11.0	1.9	14.3	.79	.45	.36
7	30	291	1.1	15.8	5.7	14.7	.74	.42	.34
8	30	309	3.3	16.0	9.2	19.2	.87	.55	.48
9	30	332	−0.0	14.8	4.7	17.5	.85	.50	.53
10	30	365	2.5	16.0	9.0	14.7	.75	.37	.45
11	30	424	2.1	12.5	8.3	13.6	.69	.45	.68
Overall	330	245	−1.1	15.3	4.1	15.9	.78	.42	.43

Note. N = 330. MEMSEN = Memory for Sentences; MEMWRD = Memory for Words.

A more complex longitudinal structural model can be written as

$$Y[t]_n = I_n + B[t] \times G_n + U[t]_n, \qquad (2)$$

which is the same as the previous equation but includes unobserved factor scores G, representing a growth score, and factor loadings $B[t]$, representing growth coefficients. The G is constant over time for each person, and the $B[t]$ are constant for the group but can change over time, so the product term $B[t] \times G_n$ allows a different impact on the outcome score $Y[t]$ at each occasion t. In other terms, G is considered as an individual slope, or trait-change score that defines the way the person changes over time and $B[t]$ is interpreted as the shape of the group growth curve. Equation 2 has previously been termed a latent growth model (following McArdle, 1988, 1989; McArdle & Epstein, 1987; McArdle & Hamagami, 1992; Meredith & Tisak, 1990).

This SEM approach allows a variety of other mod-

Table 3
Sample Statistics for 11 Monthly Time-Lag Groups on Woodcock–Johnson—Revised Reading Composite Scores (X = [LWIDNT + PSGCMP]/2)

Sample group	N	Time-lag (mean days)	Time 0 $M_{x[0]}$	Time 0 $D_{x[0]}$	Time t $M_{x[t]}$	Time t $D_{x[t]}$	Over time $R_{x[0,t]}$	Over time $R_{x[t],y[t]}$	Over time $R_{x[0],y[t]}$
1	30	68	−9.3	15.3	−8.2	16.2	.89	.54	.47
2	30	91	−3.4	14.5	0.6	12.9	.70	.19	.23
3	30	106	−6.8	13.1	−3.8	13.5	.75	.22	.26
4	30	139	2.7	16.2	7.0	13.8	.84	.25	.36
5	30	229	−0.4	20.8	3.2	18.3	.85	.44	.56
6	30	268	−5.4	14.0	1.1	13.1	.73	.63	.60
7	30	291	4.4	16.5	11.6	14.3	.70	.29	.16
8	30	309	7.8	18.2	17.2	21.3	.78	.58	.52
9	30	332	3.7	16.9	11.4	18.2	.87	.37	.46
10	30	365	−0.1	16.7	12.5	19.3	.72	.37	.44
11	30	424	4.4	19.4	15.5	21.7	.74	.36	.52
Overall	330	245	−0.2	17.1	6.6	18.7	.75	.46	.47

Note. N = 330. LWIDNT = Letter–Word Identification; PSGCMP = Passage Comprehension.

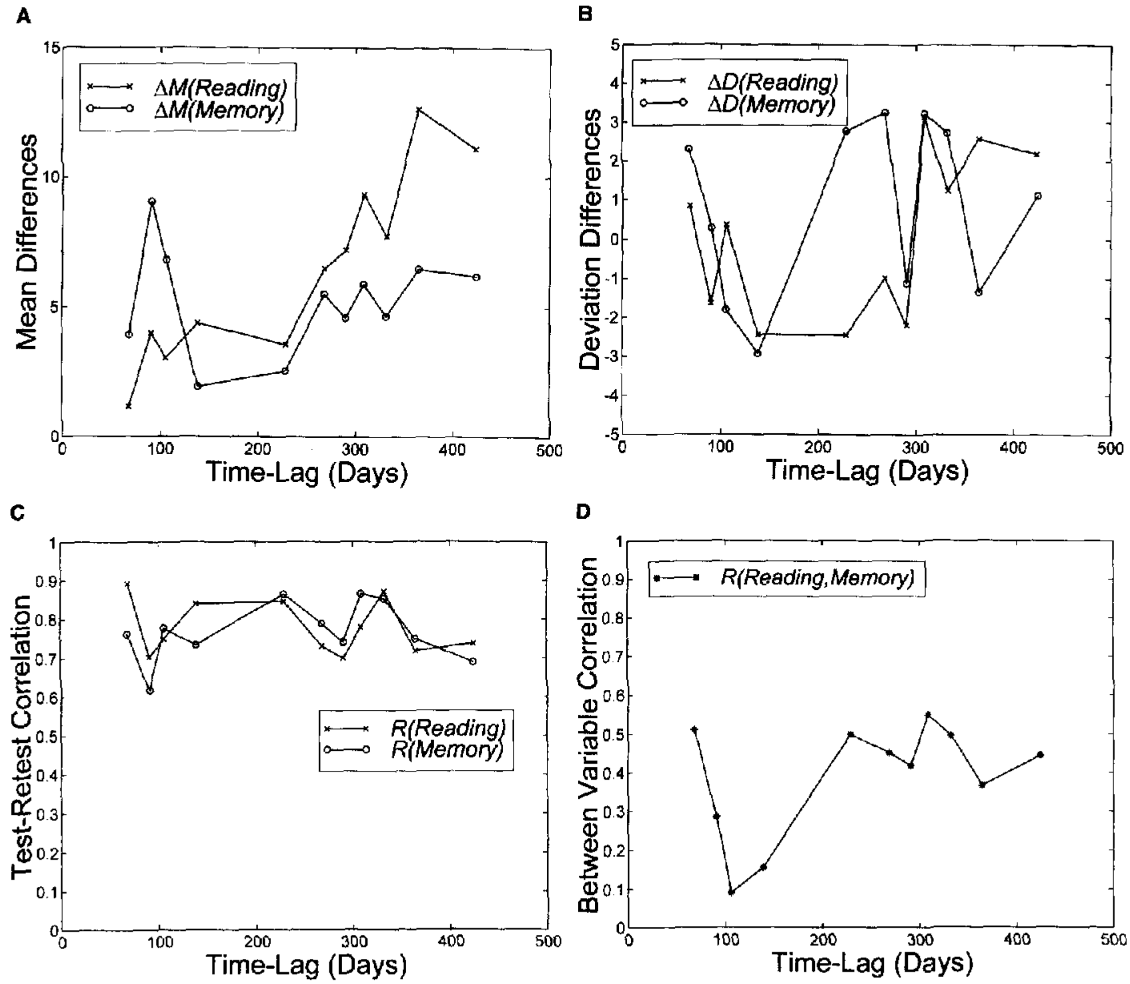

Figure 4. Observed statistics from Study 2: Memory and Reading composites for Woodcock–Johnson—Revised monthly time-lag groups. Panel A: Mean (M) changes over time. Panel B: Deviation (D) changes over time. Panel C: Test–retest correlations (R). Panel D: Between-variable correlations.

els of components of change in a univariate time-series. We can next write

$$Y[t]_n = I_n + B[t] \times G_n + A[t] \times P_n + U[t]_n, \quad (3)$$

with the additional factor score P and factor loadings $A[t]$. This model now includes multiple latent growth curves, and such models have recently been discussed by Meredith and Tisak (1990) and McArdle and Anderson (1990), among others. In this case, we add further restrictions to the loadings $A[t]$ so that P can reflect a practice or testing-effect score.

A few additional expected values ($\mathscr{E}$) are needed to express the means and covariances of the latent components. In the initial models we include non-zero means for the three common components (M_i, M_g, and M_p), non-zero variances for four components (V_i, V_g, V_p, and V_u), and at least one non-zero covariance (C_{ig}) between the initial level and growth common factor components (for a more formal expression, see the *Latent Means, Covariance, and Common Factor Notation* section in the Appendix). The non-zero covariance C_{ig} reflects the possibility that the initial level score is correlated with the growth score (e.g., Rogosa & Willett, 1985; Willett, 1988). An equivalent model can be written with this covariance estimated as a regression coefficient (as in Tucker, Damarin, & Messick, 1966). More critically, we have also assumed that the practice factor P is not correlated with either the initial level I or the growth G. These restrictions lead to a unique identification of parameters related to the P component, but such assumptions may be altered later as needed.

A Summary Path Diagram

A latent growth model of a univariate time-series is presented as a path diagram in Figure 5 (as in McArdle, 1988; McArdle & Hamagami, 1991). Following current traditions, we represent the observed variables as squares, the unobserved variables as circles, the regression coefficients as one-headed arrows, and the covariance terms as two-headed arrows.

One atypical feature of this graphic notation is the representation of all variance terms as two-headed arrows attached to the specific variables. Another unusual feature of this diagram is that the unit constant is included as a triangle, and the latent variable means (M_i, M_g, and M_p) are all represented in this diagram as the regression coefficient of a variable regressed on the constant. In this way, this path diagram explicitly includes all parameters needed to write all model matrices and expectations for the means and covariances (see McArdle & Boker, 1990).

The use of circles within squares in Figure 5 is also unusual, but it is a shorthand way of indicating the possible presence or absence of a measured variable (after McArdle, 1994). Following the data layout of Figure 2, the variable Y is always measured at the initial occasion ($Y[0]$) but may or may not be measured at each of the other time-lags ($Y[1]$, $Y[2]$, $Y[3]$, and $Y[4]$). In this notation we assume only one set of longitudinal model parameters but all measurements are not made on all occasions of interest.

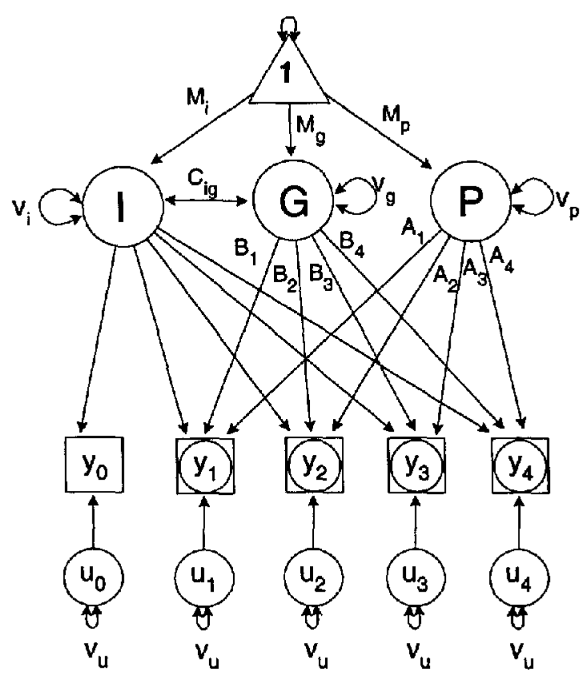

Figure 5. A latent growth model for univariate time-lag data. Square = observed variable; circle = unobserved variable; triangle = the unit constant; one-headed arrow = a unit-valued regression coefficient; two-headed arrow = a variance or covariance term.

Defining Patterns of Change

The data collected reflect a specific time series so several parameters describe patterns of change over time. Perhaps most critical here are the common factor loadings of the unknown coefficients $B[t]$ and $A[t]$. Whenever possible, we like to estimate separately the functional relationships over time ($B[t]$ and $A[t]$) as well as the means (M), deviations (D), and correlations (R) for all latent components (I, G, P, and $U[t]$). Estimation requires consideration of a variety of further substantive and mathematical model restrictions. The key questions now become: "How do we formalize an effect of growth or maturation?" "How do we formalize an effect of practice or training?" and "How do we distinguish growth effects from practice effects?"

To deal with these patterns, we first reexpress the model using standard factor analysis notation. In a model for, say, $T = 5$ occasions we can write

$$\begin{bmatrix} Y[0]_n \\ Y[1]_n \\ Y[2]_n \\ Y[3]_n \\ Y[4]_n \end{bmatrix} = \begin{bmatrix} 1 & B[0] & A[0] \\ 1 & B[1] & A[1] \\ 1 & B[2] & A[2] \\ 1 & B[3] & A[3] \\ 1 & B[4] & A[4] \end{bmatrix} [I_n\ G_n\ P_n] + \begin{bmatrix} U[0]_n \\ U[1]_n \\ U[2]_n \\ U[3]_n \\ U[4]_n \end{bmatrix}, \tag{4}$$

or, even more compactly for all persons N, as

$$\mathbf{Y} = \mathbf{L}\ \mathbf{Q} + \mathbf{U}, \tag{5}$$

where $\mathbf{L}$ is a ($T \times 3$) matrix of common factor loadings, $\mathbf{Q} = [\mathbf{I}, \mathbf{G}, \mathbf{P}]$ is a ($3 \times N$) vector of common factor scores, and $\mathbf{U} = \mathbf{U[t]}$ is a ($5 \times N$) vector of independent unique scores. In a similar matrix fashion, all means, variances, and covariances among the latent variable scores $\mathbf{Q}$ can be represented as average cross-products or moments matrix ($\mathbf{M_{qq}}$; see Browne & Arminger, 1995; McArdle, 1988; Meredith & Tisak, 1990), and these expectations can be combined to generate expectations about the observed cross-products matrix ($\mathbf{M}_{yy}$; for further details, see the Appendix).

Given this factor notation, we can now consider some alternatives based solely on restrictions of the factor loading parameters. First we consider some simple models where the loadings are fixed at some a priori value. In an initial model, we could require all $B[t] = 0$ and $A[t] = 0$ and simply write

$$\mathbf{L}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{6}$$

This matrix representation is consistent with the substantive definition of a "no growth and no practice" model. By combining these loadings $\mathbf{L}_0$ with other model matrices we end up with a highly restrictive set of model expectations (labelled $\mathscr{E}_0$ here) with no change over time.

In an alternative model we could allow "linear growth and no practice" by setting the $B[t] = t$ and the $A[t] = 0$. We write the model ($\mathscr{E}_1$) with

$$\mathbf{L}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \end{bmatrix}. \tag{7}$$

The use of this $\mathbf{L}_1$ matrix in a model allows components for both the initial level and a simple linear increase in growth. Notice that the first loading on the second component $B[0] = 0$, implying that no growth has occurred at time $t = 0$, and this helps create a mathematical separation between the first and second common factors. As we show later, this kind of model leads to predictions of increases in variances and covariances over time, but the changes in the means can be either linearly positive or linearly negative (i.e., negative growth).

More complex models can be stated using the same approach, and certain kinds of practice effects can be isolated. For example, a model with "no growth but exponential practice decay" can be formalized by setting the $B[t] = 0$ but allowing the $A[t] = e^{-\pi(t-1)}$. If we define $\pi = .2$, we can write a model ($\mathscr{E}_2$) with

$$\mathbf{L}_2 = \begin{bmatrix} 1 & 0 & .000 \\ 1 & 0 & 1.000 \\ 1 & 0 & .819 \\ 1 & 0 & .670 \\ 1 & 0 & .549 \end{bmatrix}. \tag{8}$$

In this form, the third component (P) reflects a decreasing function over time. The use of this loading matrix in a model allows components for both the initial level and the practice effects where, say, the loss is initiated at the second time point and this loss is then compounded over time. This kind of a model suggests decreasing variances and covariance over time, with the means over time following the same exponential patterning, either down or up. Incidentally, models with exponential losses or gains can be written to begin at the initial time point ($t = 0$), but this will be considered a negative growth process rather than a practice effect (see Jones, 1962; McArdle & Hamagami, 1996; Vinsonhaler & Meredith, 1966).

Other alternative models can be written to allow a mixture of both growth and practice components. These models are generally hard to estimate without a clear separation of the two developmental components, and we do not deal with all these issues here. However, one potentially useful alternative includes both "linear growth and constant practice." In model ($\mathscr{E}_3$) we define $B[t] = t$ but $A[t] = 1$ for all time points, and write

$$\mathbf{L}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 4 & 1 \end{bmatrix}. \tag{9}$$

Including these loadings in a model permits some examination of the parameters for all three common factor components I, G, and P of Figure 5.

In principle, it may be useful to fit more complex versions of change hypotheses, including models where the $B[t]$ or $A[t]$ are estimated from the available data. In model $\mathscr{E}_4$, we write

$$\mathbf{L}_4 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & B[2] & 0 \\ 1 & B[3] & 0 \\ 1 & B[4] & 0 \end{bmatrix}, \tag{10}$$

where the $B[1] = 1$ for identification purposes but the other three $B[t]$ coefficients are estimated from the data. The resulting coefficients ($T - 2$) can be used to define a flexible latent curve for the growth component. This kind of model has recently been discussed by, among others, McArdle (1988) and Meredith and Tisak (1990). This factor analytic logic can also be applied to the $A[t]$ coefficients as well. A variety of more complex growth models are possible but not used here (see Browne & DuToit, 1992; McArdle & Hamagami, 1996; see the *More Advanced Growth Functions* section in the Appendix).

Time-Lag Expectations and Estimation

Plotting Some Time-Lag Model Expectations

Some properties of the theoretical models can be understood in terms of the statistical observations generated. This requires a further translation of the linear equations (for Y) into a set of statistical expectations ($\mathscr{E}$) for the means and covariances of all observed measures over all occasions. These expectations can be formed algebraically or from the popular path analysis tracing rules (see McArdle & Boker, 1990; Wright, 1982); they will be compared with the

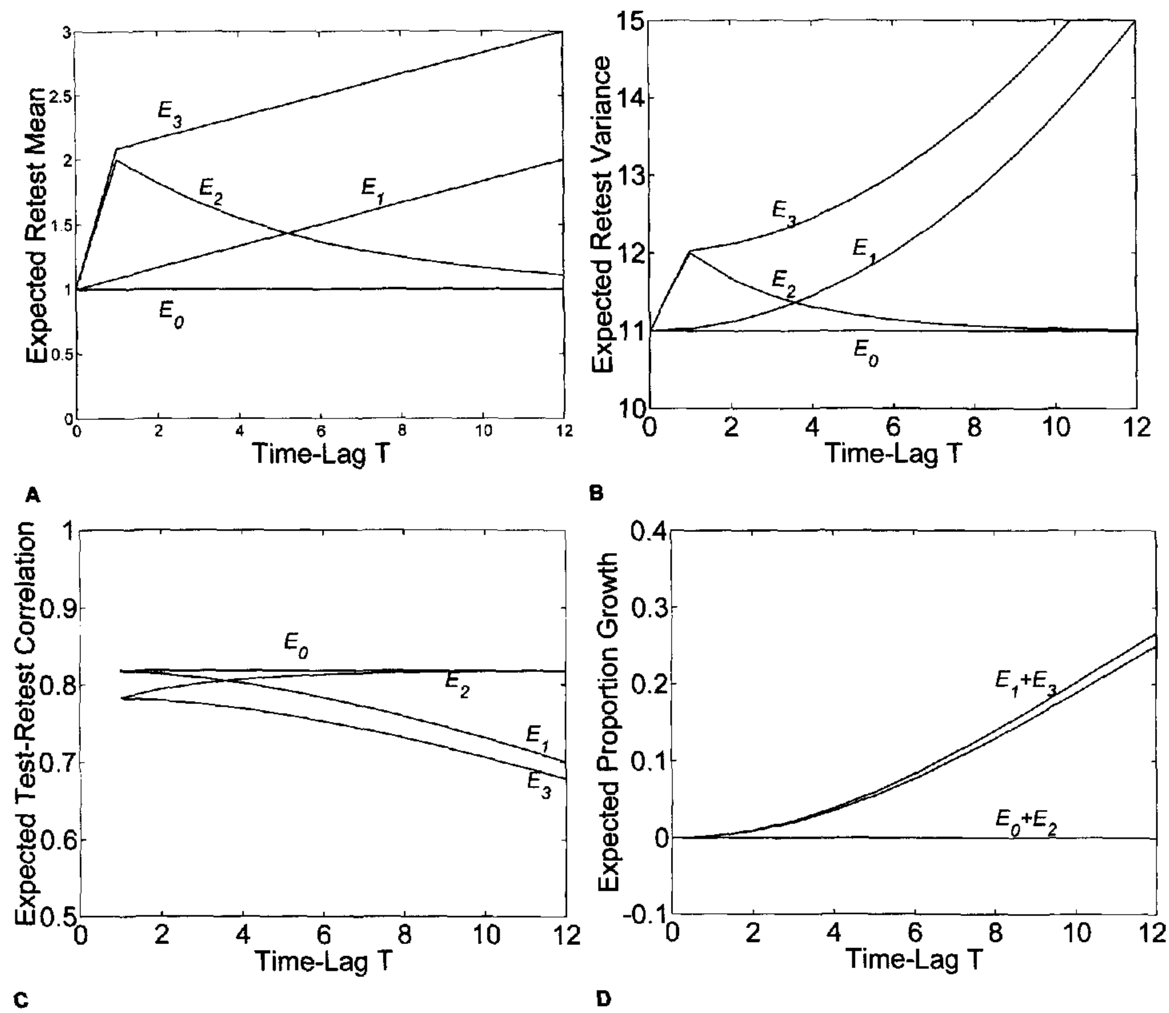

Figure 6. Theoretical time-lag characteristics of the observed statistics for four models. Panel A: Mean changes over time. Panel B: Variance changes over time. Panel C: Test–retest correlation over time. Panel D: Proportion of variance of growth. E_0 = no-change model; E_1 = linear growth model; E_2 = exponentially decreasing practice model; E_3 = linear growth with practice shift model.

observed statistics to form the optimal parameters, and they can be theoretically informative as well.

The plots of Figure 6 illustrate some basic time-lag principles. Here we show the theoretical trajectory over time for some of the expected time-lag statistics for a single variable. In each plot here we use the four factor-loading patterns of Equations 6 to 9 with identical parameters for latent means and covariance (for numerical details, see the *Time-Lag Model Mathematical Expectations* section in the Appendix). In model $\mathscr{E}_0$ we define all growth and practice terms to be zero, so this model is termed "no changes." Model $\mathscr{E}_1$ is the "linear growth" model. Model $\mathscr{E}_2$ is the "exponentially decreasing practice" model. Model $\mathscr{E}_3$ is the "linear growth with practice shift" model. The algebraic basis of each plot of Figure 6 is based on the model of Figure 5, and these are presented in detail next.

Expectations About the Means

Using standard rules of statistical expectation we can write all univariate means as

$$\mathscr{E}\{M_{y[t]}\} = \mathrm{M}_i + B[t] \times \mathrm{M}_g + A[t] \times M_p, \quad (11)$$

so the expected means $M_{y[t]}$ are a linear function of the latent means M_i, M_g, and M_p, and the coefficients $B[t]$ and $A[t]$. If we further assume that $B[0] = 0$ and $A[0] = 0$, we can simplify this expression and write

$$\mathscr{E}\{M_{y[0]}\} = M_i,$$

$$\mathscr{E}\{M_{y[t]} - M_{y[0]}\} = B[t] \times M_g + A[t] \times M_p. \quad (12)$$

These equations imply that the initial mean is based on a single parameter (M_i) and the mean changes over time-lags ($M_{y[t]} - M_{y[0]}$) are dependent on the factor loadings and factor means.

Figure 6A is a display of the means over time implied by the four models, $\mathscr{E}_0$ to $\mathscr{E}_3$, and here the four models are easily differentiated. The two linear patterns ($\mathscr{E}_1$ and $\mathscr{E}_3$) are much different from the decreasing practice ($\mathscr{E}_2$) or the no-change ($\mathscr{E}_0$) model, and all patterns over time depend on the factor loadings.

Expectations About the Variances

The expectations of the variances over time are given as

$$\mathscr{E}\{V[t]\} = V_i + B[t]^2 \times V_g + 2 \times B[t] \times C_{ig} + A[t]^2 \times V_p + V_u, \quad (13)$$

which seems more complex than the corresponding expectations for the means. If we further assume that $B[0] = 0$ and $A[0] = 0$, then these expectations can be simplified and written as

$$\mathscr{E}\{V_{y[0]}\} = V_i + V_u,$$

and

$$\mathscr{E}\{V_{y[t]} - V_{y[0]}\} = B[t]^2 \times V_g + 2 \times B[t] \times C_{ig} + A[t]^2 \times V_p, \quad (14)$$

so the variance changes over time are dependent on the factor loadings and factor variances.

As shown in Figure 6B, these variance expectations exhibit a general pattern of increases (and decreases) over time that are similar to the means squared (i.e., $\mathscr{E}\{M_{y[t]}\}^2$). A plot of the expected deviations (i.e., $\mathscr{E}\{D_{y[t]}\} = \sqrt{\mathscr{E}\{V_{y[t]}\}}$) would look very similar to the plot of the expected means.

Expectations About the Covariances

The expectations of the covariances over time can be written as

$$\mathscr{E}\{C_{y[t,t+k]}\} = V_i + B[t] \times V_g \times B[t+k] + A[t] \times V_p \times A[t+k] + B[t] \times C_{ig} + C_{ig} \times B[t+k]. \quad (15)$$

Each term here can be seen as a separate tracing in the path diagram, but these are complex and different for each pair of occasions t and $t + k$. These equations become still more complex if we assume additional non-zero correlations among all model components.

In the two-occasion test–retest data (e.g., Figure 2) we again focus on the initial occasion of measurement ($t = 0$) where we measure all participants. Because of the restrictions of $B[0] = 0$ and $A[0] = 0$, the key covariance expectation can be written more simply as

$$\mathscr{E}\{C_{y[0,t]}\} = V_i + B[t] \times C_{ig}, \quad (16)$$

and if either $B[t] = 0$ or $C_{ig} = 0$, then $C_{y[0,t]} = V_i$ for all time-lags.

By combining the previous equations we can write the standard formula for the test–retest correlation in terms of model parameters as

$$\mathscr{E}\{R_{Y[0,t]}\} = \frac{\mathscr{E}\{C_{y[0,t]}\}}{\sqrt{\mathscr{E}\,[V_{y[0]}\}}\sqrt{\mathscr{E}\{V_{y[t]}\}}} = \frac{V_i + B[t] \times C_{ig}}{\sqrt{V_i + V_u}\sqrt{V_i + B[t]^2 \times V_g + A[t]^2 \times V_p + 2 \times B[t] \times C_{ig} + V_u}}. \quad (17)$$

This equation does not directly reflect a path tracing, and it remains complex because of the variance term at time t. But this expression does highlight an important property of the time-lag model: The value of the test–retest correlation depends on the time-lag considered.

Figure 6C is a plot of these test–retest correlations over time calculated from the covariance and variance terms. The no-changes model ($\mathscr{E}_0$) has the same test–retest correlation at any time-lag whereas the other three change over time. The linear growth models ($\mathscr{E}_1$ and $\mathscr{E}_3$) both show substantially decreasing correlations over time-lags and this is due to the increasing growth variance over time. In contrast, the model of decreasing practice ($\mathscr{E}_2$) shows an increasing correlation over time is due to the eventual elimination of practice variance (e.g., Jones, 1962).

Expectations About Variance Proportions

We can also formalize some developmental components defined earlier. At each occasion t we can decompose the variance into standardized proportions (or ratios) by writing

$$V^*_{f[t]} = \frac{V_i + (B[t]^2 \times V_g) + (2 \times B[t] \times C_{ig})}{V_{y[t]}},$$

$$V^*_{p[t]} = \frac{A[t]^2 \times V_p}{V_{y[t]}}$$

and

$$V^*_{u[t]} = \frac{V_{u[t]}}{V_{y[t]}}, \tag{18}$$

where, by definition, the sum $V^*_{f[t]} + V^*_{p[t]} + V^*_{u[t]} = 1$ for any time t. Because these components can change over time, it may be useful to consider additional indices of growth and change. These might include changes in the raw deviations (e.g., $\Delta D_{f[t]} = D_{f[t]} - D_{f[0]}$), the raw variance (e.g., $\Delta V_{f[t]} = V_{f[t]} - V_{f[0]}$), or even in the standardized variance (e.g., $\Delta V^*_{f[t]} = V^*_{f[t]} - V^*_{f[0]}$) terms. Of course, any substantive interpretation of these kind of growth terms requires a meaningful starting point (at $t = 0$).

Figure 6D is a plot of the theoretical proportion of factor growth variance $V^*_f[t]$ from the four models, and only two patterns emerge. In models without growth terms ($\mathscr{E}_0$ and $\mathscr{E}_2$) the factor growth remains at the same zero level over time. In models with linear growth terms ($\mathscr{E}_1$ and $\mathscr{E}_3$) the factor growth shows increases over time. In these last models, the initial factor variance remains intact, but the factor scores have changed over time and these increasing proportions highlight this growth.

Statistical Estimation With Incomplete Data

There are many ways to use SEM to analyze longitudinal time-lag data. In longitudinal data with multiple time points, the model expectations could be applied to all pairs of occasions, $[T \times (T + 1)]/2$. If we measured, say N participants at $T = 8$ occasions of measurement we would have 44 summary statistics—eight means, eight variances, and 28 correlations. Many different models can be fitted from summary matrices of mean and covariance structures (for details, see Browne & Arminger, 1995; Horn & McArdle, 1980; McArdle, 1988, 1994; Meredith & Tisak, 1990).

The incomplete time-lag data creates several complex issues dealing with different statistics and different sample sizes. That is, from eight test–retest time-lag groups we obtain a total of 40 summary statistics—two observed means ($M_{y[0]}$ and $M_{y[t]}$), two observed variances ($V_{y[0]}$ and $V_{y[t]}$), and one observed covariance ($C_{y[0,t]}$) for each of eight independent groups. Each statistic has a potentially different (and smaller) sized sample $N[t]$. Although the information about the full (pairwise) covariance matrix is largely incomplete, we can still use the time-lag model expectations to estimate the model parameters.

The critical structural expectations for the test–retest time-lag model (i.e., Equations 12, 14, and 16) lead to at least one way we could create parameter estimates of all model components directly from the summary statistics. In one approach, a reasonable statistical estimate of the expected value of the mean level (M_i) can be obtained as the average of the means at the first time point for all scores. Similarly, the variance of the initial level (V_i) can be obtained as the average of the covariances over time-lag groups. The slope components (M_g and V_g) can be obtained by first creating averages of these statistics and then calculating $B[t]$ weighted differences; the practice components (M_p and V_p) are the intercepts in the resulting slope equations of these statistics.

This practical approach to estimation has a few problems: (a) It does not easily account for sample size differences in the summary statistics. (b) It does not provide a measure of misfit between expectations and observations. (c) It does not provide a measure of the statistical characteristics of the final parameter estimates (i.e., standard errors). (d) It would be much more complex in the presence of correlated components. (e) It can only be calculated for "fixed" $B[t]$ parameters. So, although these practical calculations can provide good initial estimates, they are neither efficient nor general solutions to this time-lag problem.

These statistical considerations suggest we use a more advanced approach to model estimation and testing, and we use statistical theory based on SEM for incomplete or missing data (Allison, 1987; Horn & McArdle, 1980; Kiiveri, 1987; Little & Rubin, 1987; McArdle, 1994; McArdle & Anderson, 1990; McArdle & Hamagami, 1991, 1996; Muthen, Kaplan, & Hollis, 1987; see Appendix). The SEM analyses we present next are based on maximum-likelihood estimation (MLE) of the means and covariances to account for the incomplete patterns and different sample sizes, but other weighted fitting functions (e.g., GLS) could be used as well.

Estimation Using Standard SEM Software

The aggregation of individuals into time-lag groups allows us to analyze a variety of longitudinal models using any available SEM software that permits a multiple group calculation (e.g., LISREL-8 by Jöreskog & Sörbom, 1993; Mx by Neale, 1993; also see McDonald, 1980). The key feature of this programming is that the overall model parameters remain invariant but they are deployed systematically among the different time-lag groups. More complete details on the required computer programming are presented in the Appendix.

One useful byproduct of MLE is the calculation of a likelihood ratio test (LRT) statistic for the evaluation of goodness-of-fit. These LRT indices and their differences (dLRT) can be compared to a chi-squared distribution with degrees of freedom based on the number of summary statistics minus the number of model parameters. Other useful byproducts of MLEs include the calculation of standard errors, confidence boundaries, and other indices of goodness-of-fit. In recent work, Browne and Cudeck (1993) suggested an index of "close fit" to the data, based on the root mean square error of approximation (i.e., RMSEA < .05) and this overall criterion of fit will be used here. Calculation of statistical power for incomplete data designs is also possible using MLE techniques (see McArdle, 1994; McArdle & Hamagami, 1992). These and other useful properties of MLE are often based on assumptions of multivariate normality of the model residuals (see Browne & Arminger, 1995), so other fitting functions may be needed.

Time-Lag Results

Study 1: Results for the Daily Univariate WJ–R Data

We fit a variety of longitudinal models to the daily MEMNAM statistics of Table 1 (i.e., eight groups, each with one correlation, two deviations, and two means). The results for five alternative models are listed in Table 4.

The first column of parameters (labeled $\mathscr{E}_0$) in Table 4 is based on a "no growth and no practice" or "initial level only" model fitted to the summary statistics of Table 1. To fit this model we estimated only three parameters (i.e., M_i, V_i, V_u) using the $\mathbf{L}_0$ matrix (defined in Equation 6). The estimated parameters are all significantly different from zero, but the goodness-of-fit obtained is poor: LRT = 961 on df = 37; RMSEA = .135.

The second column ($\mathscr{E}_1$) lists the results of a "linear growth and no practice" model fitted to the same data. In this model we estimate six parameters (i.e., M_i, V_i, V_u, M_g, V_g, C_{ig}) using the $\mathbf{L}_1$ matrix (see Equation 7). Here we obtained small but significant negative parameters for the means (e.g., $M_g = -2.5$) and a nonsignificant growth variance, V_g. This indicates the decline in the means is not similar to changes in the covariances, but it also shows a negative growth pattern for the means. The goodness-of-fit obtained now is still a poor fit (LRT = 242 on df = 34; RMSEA = .067), but the addition of the growth component improves the change in fit a great deal (ΔLRT = 719 on Δdf = 3).

The third column ($\mathscr{E}_2$) gives the results of an "exponential practice and no growth" model. Here we estimated five parameters (i.e., M_i, V_i, V_u, M_p, V_p) using a set of loadings similar to $\mathbf{L}_2$ (see Equation 8) plus one extra parameter $\pi = .093$ used to form all loadings. This means that a first component is the initial level I and the second component P is a practice effect which starts at the $t = 1$ and decays exponentially over successive time points. The significant practice mean ($M_p = -8.1$) indicates an 8-point loss in scores at the time-lag of 1 day. Once again, the variance associated with this second component was very small, so this best reflects the group decline and not the individual changes. The goodness of fit obtained now is much better (LRT = 52.7 on df = 35; RMSEA = .020), and the addition of this practice component substantially improves the relative fit (ΔLRT = 908 on Δdf = 3).

The next column ($\mathscr{E}_3$) gives the results of a "linear

Table 4
Univariate Estimates for Alternative Models Fitted to Study 1 Memory-for-Names Daily Time-Lag Data

Model parameter estimated	$\mathscr{C}_0$ level only	$\mathscr{C}_1$ linear growth	$\mathscr{C}_2$ exponential practice	$\mathscr{C}_3$ linear + shift	$\mathscr{C}_4$ latent growth
Model coefficients					
Growth $B[t]$	0=	t=	0=	t=	$B[t]$
Practice $A[t]$	0=	0=	π = .093*	1=	0=
Initial M_i	−4.9*	−1.1*	.1	.1	.1
Growth M_g	0=	−2.5*	0=	−1.0*	−8.2*
Practice M_p	0=	0=	−8.1*	−6.9*	0=
Model variances and covariances					
Initial V_i	109.*	102.*	101.*	115.*	108.*
Growth V_g	0=	.2	0=	.7*	5.8*
Covar. C_{ig}	0=	−.7	0=	−4.8*	−6.0*
Unique V_u	74.0*	61.5*	56.1*	43.7*	51.5*
Practice V_p	0=	0=	.3	17.3*	0=
Goodness-of-fit indices					
Free parameters	3	6	6	8	13
Degrees of freedom	37	34	34	32	27
Likelihood ratio	961.0	242.0	52.7	44.9	46.1
Prob. perfect fit	<.01	<.01	<.02	<.06	<.02
RMSEA index	.135	.067	.020	.017	0.23
Prob. close fit	<.01	<.01	<1.0	<1.0	<1.0
Standardized variance components (assuming t = 7 for 1-week lag)					
Initial $V_f^*[0]$	.596	.624	.643	.724	.677
One week $V_f^*[7]$	.596	.624	.639	.568	.670
Practice $V_p^*[7]$	.000	.000	.006	.122	.000
Unique $V_u^*[7]$	.404	.376	.355	.309	.330
Test–Retest $R_y[0,7]$	.596	.593	.641	.542	.620

Note. This table is based on age-partialled data with N = 1,384 from Table 1 and maximum-likelihood estimates from LISREL-8 and Mx-92. An asterisk indicates parameter that is larger than 1.96 times its standard error; equal sign indicates a parameter has been fixed for identification. M_i and V_i are the result of the age-regression adjustments. The $\mathscr{C}_2$ loadings $A[t] = e^{(.093*(t-1))}$ = [0, 1, 1.10, 1.21, 1.32, 1.45, 1.59, 1.75 and 1.92]. The $\mathscr{C}_4$ loadings $B[t]$ = [0, 1, 1.08, 1.14, 1.54, 1.46, 1.52, 1.70, and 2.07]. Prob. = probability of; RMSEA = root mean square error of approximation from Browne and Cudeck (1995).

growth plus a practice shift'' model. Here we estimate all eight parameters (i.e., M_i, V_i, V_u, M_p, V_p, M_g, V_g, C_{ig}) using the matrix $\mathbf{L}_3$ (see Equation 9). The first component is the initial level I, the second component is the linear growth G, and the third component P is a second initial level which starts at the t = 1 and remains constant from that point on. The negative growth mean (M_g = −1.0) once again indicates a loss over time that accumulates linearly with increasing time-lag. In contrast, the significant practice mean (M_p = −6.9) indicates a seven-point loss in memory scores at the second time for any time-lag. The estimated variance proportions show the initial variance is large but the growth in this variance is very small. The variance associated with the practice shift component is larger (V_p = 17.3 with $V^*_{p[7]}$ = .109) and this indicates substantial individual differences in practice that is not related to the time-lags used here. On a statistical basis, model $\mathscr{C}_3$ provides an excellent fit to these Memory-for-Names data (LRT = 44.9 on df = 32; RMSEA = .017).

The final column ($\mathscr{C}_4$) gives the results of a latent basis model. Here we estimate six of the previous parameters (i.e., M_i, V_i, V_u, M_g, V_g, C_{ig}), but we also estimate seven $B[t]$ elements using a matrix similar to $\mathbf{L}_4$ (see Equation 10). This means that the second component G allows a flexible form for the growth function over time. The resulting parameters indicate a negatively decreasing function (M_g = −8.2), which has small variance (V_g = 5.8). The seven estimated $B[t]$ curve coefficients rise only slightly from lags of 1 day (t = 1) to lags of 8 days (t = 8). A good fit is obtained here (LRT = 46 on df = 27; RMSEA = .023) but the single curve model $\mathscr{C}_4$ with many parameters does not fit as well as the simpler model $\mathscr{C}_3$.

The 8 parameters of the best fitting model $\mathscr{C}_3$ yields the 40 expectations for the means, deviations, and covariances listed previously in the bottom of Table 1. The small differences between the observed values in the top of Table 1 and the expected values in the bottom of Table 1 leads to the close fit of the model $\mathscr{C}_3$.

Study 2: Results for the Monthly Univariate WJ–R Data

Similar univariate longitudinal models have been fitted to the data from the second WJ–R study. All univariate models were fitted to the 55 independent test statistics for each composite variable listed in Tables 2 and 3 (i.e., 11 groups, each with 1 correlation, 2 deviations, and 1 mean). For brevity here we only discuss results from models using loadings with a linear growth $B[t]$ and an initial practice $A[t] = 1$ (i.e., of type $\mathbf{L}_3$). The linear growth coefficient matrix $B[t]$ was scaled (at t) so that one unit in this metric equals 1 year of time-lag. Results from the best fitting models are presented in the first two columns of Table 5.

Model $\mathscr{C}_5$ of Table 5 gives estimates for a "no growth or practice only" model fitted to the data on the WJ–R Short Term Memory score (of Table 2). In this initial univariate model we have fixed the uniqueness at a value based on the previously published internal consistency (i.e., $V_u = 25.8$; the Appendix's section on WJ–R data). This no-growth model (i.e., $M_g = V_g = C_{ig} = 0$) yielded a significant practice mean ($M_p = 5.2$) indicating a 5-point gain in Memory scores just for having taken a retest at any time-lag. We also obtained a large initial level variance ($V_i = 199$), a smaller state variance ($V_s = 20.5$), a nonsignificant practice variance ($V_p = 20.9$), and a close fit (RMSEA $= .038$) to the Short Term Memory data.

Model $\mathscr{C}_6$ in Table 5 gives the result of a "no practice or growth only" model fitted to the data on the WJ–R Broad Reading score of Table 3. We again fixed the factor loading (at $H = 1$) and the uniqueness at a value specified by the internal consistency (i.e., $V_u = 15.4$). This no-practice model (i.e., $M_p = V_p = 0$), yielded a significant linear growth mean ($M_g = 10.2$), indicating a 10-point gain in reading scores for every 1-year interval of time. We also obtained a large initial level variance ($V_i = 264$), a large growth variance ($V_g = 98.7$), a small but significant state variance ($V_s = 22.3$), and a close fit (RMSEA $= .040$) to the Reading data.

Summary of Univariate Results

In Study 1, we expected the memory losses would increase with longer time between test and retest in the daily Memory-for-Names task. We also expected substantial individual differences in these memory losses, perhaps as a single functional form and possibly correlated with the initial level of memory. This final model chosen (see $\mathscr{C}_3$ of Table 4) yields some interesting substantive information about these hypotheses. The decreasing pattern in the means ($M_{y[t]}$) shows the group has an initial 8-point loss for 1 day, a 9-point loss for 2 days, a 10-point loss for 3 days, and so on. However, the large initial-level variance and very small growth variance suggests this decline reflects only the group means and is not related to a single source of systematic individual differences indicated by the covariances. Thus, a general decline in memory over time was found, but after we hold constant the contamination due to simple practice effects, our initial hypothesis about a single function of memory loss was not substantiated.

In Study 2, the univariate results from the monthly time-lag data on Reading and Memory (Study 2) yielded different substantive results. The Short-Term Memory scores were fitted well by a "no-growth" model (see $\mathscr{C}_5$ of Table 5), and this shows substantial practice effects in mean and variance over any monthly time-lag. This pattern would be expected of a psychological variable that has lots of short-term variation or has already reached some peak level. In contrast, the reading scores were fitted well by a "no-practice" model (see $\mathscr{C}_6$ of Table 5), and this shows a substantial linear growth pattern over the entire yearly period of time-lag between tests. This pattern would be expected of a psychological variable that is undergoing growth and has not already reached some peak level.

These univariate models demonstrate our general approach to modeling, but the substantive results can be enhanced and clarified in several ways. It would be informative to make direct comparisons between variables, especially comparisons based on individual differences. Also, we would like to be able to estimate the unique variance (V_u was fixed above) and also make some estimate of state variance (V_s was not estimated above). These substantive issues lead us to consider more complex multivariate data and models.

Table 5
Univariate and Bivariate Estimates From Selected Models Fitted to the Study 2 Monthly Woodcock–Johnson—Revised Time-Lag Data (of Table 2)

	Univariate				Bivariate			
	$\mathscr{C}_5$ memory	$\mathscr{C}_6$ reading	$\mathscr{C}_7$ general		$\mathscr{C}_8$ memory		$\mathscr{C}_9$ reading	
Model parameter estimated	Y_{1+2}	Y_{3+4}	Y_{1+2}	Y_{3+4}	Y_1	Y_2	Y_3	Y_4
Test-specific coefficients								
Factor L_w	1=	1=	1=	1.92*	1=	.976*	1=	.727*
Practice $A[t]_w$	1=	1=	1=	1=	1=	1=	1=	1=
Practice M_{pw}	5.2*	0=	1.5	–.8	6.1*	–.9	0=	0=
Intercept M_{iw}	–1.1	–.3	–1.1	–.2	–.9	–1.4	3.0*	–3.4*
Test-specific variance-covariances								
Unique V_{uw}	25.8=	15.4=	182.*	53.8*	61.7*	178.*	69.7*	164*
Practice V_{pw}	20.9	0=	0<	0<	0<	0<	0=	0=
Trait-specific coefficients								
Growth $B_{[t]}$	t/12	t/12	t/12		t/12		t/12	
Growth M_g	0=	10.2*	5.8*		0=		11.4*	
Trait-specific variance-covariances								
Initial V_i	199.*	264.*	66.1*		200.*		340.*	
Growth V_g	0=	98.7*	11.4		0=		120.*	
Covar. C_{ig}	0=	–29.7	–.7		0=		–41.1	
State V_s	20.5*	22.3*	0<		.2		0<	
Goodness-of-fit indices								
Free parameters	5	6	14		11		10	
Degrees of freedom	50	49	140		143		144	
Like. ratio	73.1	73.5	386.0		219.0		332.0	
Prob. perfect fit	<.02	<.02	<.01		<.01		<.01	
RMSEA Index	.038	.040	.077		.041		.064	
Prob. close fit	<.85	<.82	<.01		<.92		<.01	

Notes. This table is based on age-partialled data with N = 330 from Tables 2–3 and maximum-likelihood estimates from LISREL-8 and Mx-92. An asterisk indicates parameter that is larger than 1.96 times its standard error; equal sign indicates a parameter has been fixed for model identification; less-than sign (<) indicates a parameter that remained on a boundary. Basis B are fixed equal to linear trend with 1 year proportion (i.e., t/12). Y_1 = MEMSEN; Y_2 = MEMWRD; Y_3 = LWIDNT; Y_4 = PSGCMP.

A Multivariate Time-Lag Extension

Including a Multivariate Measurement Model

There are many ways to expand the models of the previous sections. To include a complete empirical separation of all developmental concepts discussed earlier, we need to expand to a multivariate form. One way to do this is to write a factor measurement model for the observed scores as

$$Y_{w,n} = H_w \times F_n + U_{w,n}, \tag{19}$$

where, for each separate variable Y_w, w is a numerical index (with scores Y_1, Y_2, Y_3, etc.), the coefficients H_w are common factor loadings, the F is the unobserved common factor or true score, and the U_w is the unique factor score. As in standard factor analytic treatments, this unique score is theoretically the sum of an error-of-measurement score and a specific factor score. One way to add a time-lag to this model is to write

$$Y[t]_{w,n} = H_w \times F[t]_n + U[t]_{w,n}, \tag{20}$$

where the factor scores $F[t]$ are assumed to change with time but the factor pattern H is assumed to have factorial invariance over time (Horn & McArdle, 1980, 1992; McArdle, 1988; Meredith & Tisak, 1990).

Let us further assume that the previous growth model can be directly applied to the factor scores by writing

$$F[t]_n = I_n + B[t] \times G_n + S[t]_n. \tag{21}$$

This model uses the same notation as in the earlier models for growth coefficients $B[t]$ and growth scores G, but here we represent the growth in the unobserved common factor scores. In this model of the common factor scores we also include a second-level unique score termed $S[t]$, which is common to all variables Y_w within a time and (as with $U[t]$ before) is independent of other $S[t + l]$ across time and has zero mean. By these definitions, the $S[t]$ can be termed a *common state* or *state-fluctuation score*. In contrast to other treatments (e.g., Steyer, 1989; Steyer et al., 1990), this state variable is considered as the nongrowth or transitory component of the factor score within each time (after Horn & Little, 1966; Horn, 1972; Nesselroade, 1993).

Let us next add the practice component P discussed earlier by writing

$$Y[t]_{w,n} = H_w \times F[t]_n + A[t]_w \times P_{w,n} + U[t]_{w,n} \quad (22)$$

where, for each variable w, we add a separate practice score P_w and loading $A[t]_w$. By this definition each test has a specific practice component.

By combining the previous equations we can also write the model for observed scores as

$$Y[t]_{w,n} = H_w \times (I_n + B[t] \times G_n + S[t]_n) + A[t]_w \times p_{w,n}$$

or

$$= H_w \times I_n + H_w \times B[t] \times G_n + H_w \times S[t]_n + A[t]_w \times P_{w,n} + U[t]_{w,n}, \quad (23)$$

so the model is seen to have five separate components (I, G, S, P, U) with multiplicative coefficients. This multivariate model can also be seen as a higher order factor analysis model with first-order measurement loadings H, with second-order growth loadings $B[t]$, and with some consideration of specific practice effects $A[t]$. This kind of multivariate model allows for both stability and change components in the tests ($Y[t]_w$) and in the traits ($F[t]$) and has been termed a *curve of factor scores* model (McArdle, 1988). Specific factor components and additional multivariate features may now be added as necessary.

A Mutivariate Path Diagram

This multivariate longitudinal model is presented in the path diagram of Figure 7. We assume the same variable $Y[t]_w$ has been repeatedly measured on at least two occasions, these measured scores are an outcome of unobserved common factors $F[t]$ and have independent unique scores $U[t]_w$. The factor scores $F[t]$ are influenced by additional factors labeled initial I, growth G and state $S[t]$. The I score influences all $F[t]$ scores equally but the G score only influences the later $F[t > 0]$ scores with changing $B[t > 0]$. The independent state components $S[t]$ are seen to have an impact only on the factor score within an occasion, are uncorrelated over time, and have zero mean (i.e., no relation to the unit constant 1). Finally, the test-specific practice score Pw influences each observed test score at the later time points ($t > 0$), and has both a test specific mean (M_{pw}) and a test specific variance (V_{pw}). The means at the first occasion ($M_{yw[0]}$) are estimated from the initial level parameters M_{iw} (these are not drawn), but the means at later occasions ($M_{yw[t]}$) depend on the both the common latent growth and specific practice means.

Multivariate Expectations and Variance Components

The multivariate expectations are more complex but can be formed by a combination of the previous concepts and equations (for details, see the *Multivariate Time-Lag Expectations* section in the Appendix). The expectations for the multivariate means require the possibility of an arbitrary scaling constant or intercept M_{iw} for each variable (not drawn in Figure 7). We then write all observed score means in terms of the factor means $M_{f[t]}$ and the factor loadings H_w. Similarly, expectations for the covariances within each variable can be written in terms of the factor covariances over time ($C_{f[t,t+l]}$) of covariances within a time ($C_{j[t],k[t]}$) and covariances between different times ($C_{j[t],k[t+1]}$). In all cases above, the observed covariances can be seen as functions of the latent variable covariances.

The univariate models allow independent estimates of several useful variance components for each observed variable w. First we can define a *factor communality* ratio ($R_{fc(y[t]w)}$) to index the proportion of the initial common factor variance included in variable w at time t. Similarly we can define a *practice–retest* ratio ($R_{pr(y[t]w)}$) to index the proportion of practice variance in any observed variable Y_w at time t. Other test-specific ratios can be formed from the five components of Equation 16 in various ways.

In this multivariate framework we have an added opportunity to define ratios that are *trait-specific*. In these ratios, the denominator is the common factor variance ($V_{f[t]}$) at occasion t. A *factor stability* ratio ($R_{fs[t]}$) may be defined as the proportion of the initial level of the trait that remains in the common factor scores at time t. Likewise, a *factor-growth* ratio ($R_{fg[t]}$) may be defined as the proportion of the systematic growth or change variance which is now included in the common factor scores at occasion t. Finally, a

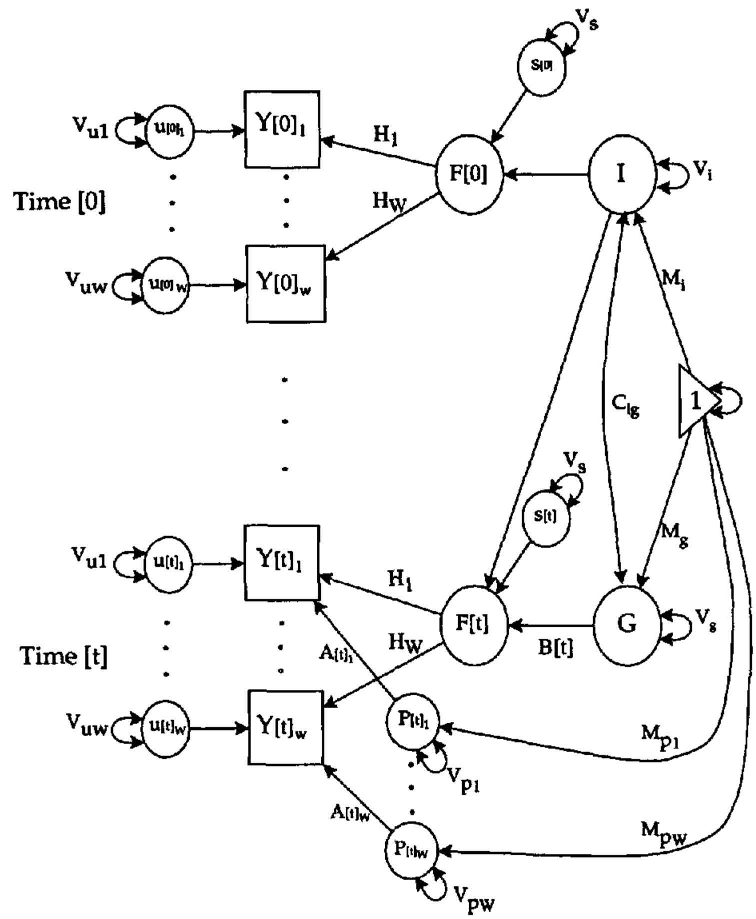

Figure 7. A latent growth path model for multivariate time-lag data. Square = observed variable; circle = unobserved variable; triangle = the unit constant; one-headed arrow = a unit-valued regression coefficient; two-headed arrow = a variance or covariance term.

state-fluctuation ratio ($R_{sf[t]}$) may be defined as the independent common factor variance in the factor scores at any time t. This last coefficient is common to all variables within an occasion and will be separate from the test-specific unique variance ($V^*_{u[t]}$). These trait-specific proportions can be written so the sum is unity within any time point t (i.e., $R_{fs[t]} + R_{fg[t]} + R_{sf[t]} = 1$) so these proportions are only useful when we have a meaningful starting point ($t = 0$).

Results from Multivariate Time-Lag Models

Results From Bivariate Factor Models

Several bivariate models were fitted to the monthly WJ–R data (Study 2) discussed earlier. These models were each based on only two variables following the path diagram of Figure 7. To identify all model parameters here we used standard factor analytic identification constraints: (a) We fixed the factor loading for one variable ($H_w = 1$). (b) We estimated the second loading and both unique variances. (c) We equated the loadings at the second occasion (i.e., factorial invariance). (d) We allowed separate intercepts for each variable (M_{iw}). (e) We equated these intercepts at the second occasion. (f) We forced all mean differences over time to be accounted for by the common factors ($F[t]$).

A single factor model $\mathscr{E}_7$ was initially fitted to all memory and reading time-lag data of Tables 2 and 3. This analysis includes 14 parameters fitted to 154 summary statistics (for 11 groups, each with 4 means, 4 variances, and 6 correlations). This bivariate model proved to be extremely cumbersome to fit and a variety of additional boundary conditions were needed to produce numerical convergence (i.e., $V_s > 0$, $V_p > 0$). The relative loadings for Memory ($H_1 = 1.00$) and

Reading (H_2 = 1.92) suggest that the Reading composite dominates this General factor, but further interpretation is not needed due to the extremely poor fit of the model (LRT = 386 on df = 140; RMSEA = .077).

As an alternative approach, we fit the bivariate model at a lower level of measurement. A Short Term Memory factor model $\mathscr{C}_8$ was fitted using the more basic scales—Memory for Sentences (MEMSEN) and Memory for Words (MEMWRD) WJ–R scales. The numerical results obtained now show about equal loadings (H_2 = .976; see Table 5) and the unique variances are larger than the univariate model estimate (V_u = 61.7, 178) indicating substantial specific factor components. The practice effects differ slightly between the two variables: the MEMSEN has a strong practice mean (M_{p1} = 6.1), the MEMWRD practice mean is almost zero, and neither variable has a significant practice variance. The common factor of these two variables has a large initial level variance (V_i = 200), and the common state variance is very close to zero in this model. The standardized factor loadings are estimated as $H[t]^*$ = [.764, .517] and, due to no growth and no practice variance, these standardized loadings are the same at all time points. This model, assuming no-growth in a common factor of Memory, provides an excellent fit (e.g., RMSEA = .041).

The Broad Reading factor model $\mathscr{C}_9$ was fitted to the other WJ–R scales—Letter-Word Identification (LWIDNT) and Passage Comprehension (PSGCMP). The numerical results show lower loadings for the second test (H_2 = .727) and large unique variances (V_{u1} = 69.7, V_{u2} = 164.). This indicates potentially important differences in the constructs measured by these two tests. Nevertheless, the common factor of these two variables has a large initial level variance (V_i = 340), corresponding large linear growth over time in both the mean growth (M_g = 11.4) and growth variance (V_g = 120), and the estimated common factor state variance is close to zero. The standardized estimates of the factor loadings are calculated as $H[0]^*$ = [.830, .523] at t = 0, and, due to the growth variance, increase to $H[7]^*$ = [.883, .597] at t = 1 year. This no-practice common factor model of Reading yields a questionable fit to these WJ–R data (e.g., RMSEA = .064) so other alternatives may be needed.

Alternative Developmental Components and Hypotheses

All previous model estimates can be recast as developmental components, and these are calculated for each variable in the columns of Table 6. In contrast to the initial univariate estimates, these bivariate calculations show the factor stability coefficients are raised, and the state-fluctuation variances are nearly zero. In comparable cases, the overall pattern of changes in the latent common factor can be seen as enhanced versions of the univariate estimates.

The models above presented only the most restrictive hypotheses about practice and growth. But a variety of alternative models can be fitted before making any firm conclusions. Table 7 presents goodness-of-fit indices for some of these models. The first row of Table 7 gives the overall fit indices (LRT and df) for a model where all parameters have been fit to each dataset. All of these initial fits are excellent except for

Table 6
Resulting Standardized Components for Monthly Test–Retest Models (Assuming B[t] = *1 for 1 Year Lag)*

	Univariate				Bivariate			
	$\mathscr{C}_5$ Memory	$\mathscr{C}_6$ Reading	$\mathscr{C}_7$ General		$\mathscr{C}_8$ Memory		$\mathscr{C}_9$ Reading	
Component calculated	Y_{1+2}	Y_{3+4}	Y_{1+2}	Y_{3+4}	Y_1	Y_2	Y_3	Y_4
Test-specific components								
Internal consistency $R_{ic}[0]_w$	.895	.949	.895	.949	.862	.794	.936	.887
Test–retest correlation $R_y[0, 1]_w$	.779	.731	.259	.765	.764	.517	.715	.447
Factor communality $V^*_{fc}[0]_w$	.895	.949	.266	.518	.765	.518	.851	.522
Practice variance $V^*_{pr}[1]_w$	.079	.000	.000	.000	.000	.000	.000	.000
Trait-specific components								
Factor stability V^*_{fs}	.906	.718	.859		.999		.792	
Factor growth $V^*_{fg}[1]$	.000	.214	.141		.001		.208	
State fluctuation $V^*_{sf}[1]$	.094	.068	.000		.000		.000	

the bivariate General factor and Reading factor model discussed above.

The second row of Table 6 gives the fit for the no-growth hypotheses (i.e., $M_g = V_g = C_{ig} = 0$) from all previous datasets. Because this second model is a nested subset of the first model we can calculate the difference in fit, and this shows a clear pattern in both univariate and bivariate models: No growth is reasonable for common Memory scores (ΔLRT = 4 on Δdf = 3) but is not reasonable for common Reading scores (ΔLRT = 56 on Δdf = 3). The next row gives results for the no-practice hypothesis (i.e., $M_{pw} = V_{pw} = 0$), and these multivariate results are less than clear: No practice seems unreasonable for Memory scores (ΔLRT = 7 on Δdf = 2) but does seem reasonable for Reading scores (ΔLRT = 3 on Δdf = 2). The last model sets all practice and growth parameters to zero. A "no changes" model shows a marked loss of fit for all cognitive data described here.

Results From Multivariate Factor Models

Two final multivariate models were fitted using all four WJ–R scales together. That is, for each of the 11 monthly groups, these analyses included eight means, eight standard deviations, and 28 correlations (for a total of 484 summary statistics; not listed here). Table 8 gives the results for a two-common factor model ($\mathscr{C}_{10}$) and a one-common factor model ($\mathscr{C}_{11}$).

Model $\mathscr{C}_{10}$ includes two common factors, a Memory factor (based on Y_1 and Y_2) and a Reading factor (based on Y_3 and Y_4). This factor model includes two sets of factor loadings, H = [1.00, .75; 1.00, .97], and two sets of trait change parameters. In addition, this model also includes covariance parameters (C_f) relating the developmental components among the factors. This is a restrictive multivariate model because the only covariances allowed are between the initial levels and growth parameters. The resulting test and trait coefficients are very similar to bivariate estimates (in $\mathscr{C}_3$ and $\mathscr{C}_9$ of Table 5), and only a few of the latent trait covariances are noteworthy. The correlation of the initial levels is $R_{i1,i2}$ = .55 (calculated from the estimated variance and covariances; $137/\sqrt{330} \times \sqrt{192}$), but the covariance of all other latent growth components is nearly zero. The goodness-of-fit of this restrictive two-factor model with 32 parameters is quite good (LRT = 744 on df = 452; RMSEA = .045).

Model $\mathscr{C}_{11}$ is based on the same data, but it includes only one common factor. This model includes three free factor loadings H = .64, .79, 1.00, and .86, and posits all individual differences in both initial level and growth can be organized by a single general factor. The model parameters for the loadings are all relatively high (H_w > .6), but all common growth variance is nearly zero. Perhaps more importantly, the goodness of fit of this 24 parameter model is no longer adequate (LRT = 1165 on df = 460; RMSEA = .069). The difference in fit between the two-common factor model $\mathscr{C}_{10}$ and this one-common factor model $\mathscr{C}_{11}$ is relatively large (ΔLRT = 421 on Δdf = 8), so we conclude that the one-factor model does not fit these data.

Summary of Multivariate Results

The results from the monthly time-lag multivariate data on Reading and Memory (Study 2) yield some interesting substantive results. First, simultaneous estimation of all parameters, including both the unique variances (V_u) and the state variance (V_s), was esti-

Table 7
A Summary of Goodness-of-Fit Indices for Alternative Test–Retest Models

	Univariate				Bivariate					
	Memory		Reading		General		Memory		Reading	
Model comparisons	χ^2	df	χ^2	df	χ^2	df	χ^2	df	χ^2	df
1. Overall	69	47	71	47	386	140	214	140	317	140
2. No growth	73	50	127	50	424	143	219	143	373	143
difference (2 − 1)	4	3	56	3	38	3	5	3	55	3
3. No practice	76	49	74	49	388	144	227	144	332	144
difference (3 − 1)	7	2	3	2	2	4	13	4	15	4
4. No change	144	52	215	52	524	147	292	147	477	14
difference (4 − 1)	75	5	144	5	138	7	88	7	160	7

mated but some components (V_p and V_s) were probably not needed here. The poor fit of the bivariate model across different variables provides some evidence that Memory and Reading composites do not exhibit the same general latent change patterns. However, the good fit of the same within each pair of variables of similar content suggests that these Reading and Memory composites do have substantial validity. Finally, the four variable models provide a direct test of the one-factor hypothesis and this strongly suggests that more than one common factor is needed to account for these Memory and Reading data.

In sum, a single factor has little construct validity here because this model does not account for both the within-time and across-time information in these cognitive abilities. Although consideration of Spearman's g remains a typical hypothesis in multivariate cross-sectional data analysis (e.g., Horn, 1988, 1991), these repeated measures analyses add considerable power to the statistical hypotheses (see McArdle & Nessel-

Table 8
A Summary of Two Alternative Multivariate Test–Retest Models Fitted to the Monthly Woodcock–Johnson—Revised Time-Lag Data

	Two-factor $\mathcal{C}_{10}$							
	Memory		Reading		One-factor $\mathcal{C}_{11}$ General			
Parameter estimated	Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
Test coefficients								
Loading H_w	1 =	.973*	1 =	.748*	.641*	.643*	1 =	.862*
Practice $A[t]_w$	1 =	1 =	1 =	1 =	1 =	1 =	1 =	1 =
Intercept M_{iw}	3.2*	−3.6*	−0.9	−1.4	−0.9	−1.4	3.2*	−3.6*
Practice M_{pw}	−0.7	−0.1	3.9*	5.6*	0.3	0.2	0.9	0.4
Test variance–covariances								
Unique V_{uw}	69*	157*	56*	178*	142*	254*	148*	134*
Practice V_{pw}	0<	0<	11	0<	13	0<	0<	35
	Memory		Reading			General		
Trait coefficients								
Growth $B_{[t]}$	t/12		t/12			t/12		
Growth M_g	0.6		12.1*			9.4*		
Trait variance–covariances								
Initial V_i	192*		330*			250*		
Growth V_g	0<		102*			0<		
Covariance C_{ig}	13		−32			18		
State V_s	0<		0.4			0<		
Covariances $C_{i1,i2}$, $C_{g1,g2}$	3		137*					
Covariances $C_{i1,g2}$, $C_{g1,i2}$	10		25*					
	Two-factor $\mathcal{C}_{10}$					One-factor $\mathcal{C}_{11}$		
Goodness-of-fit indices								
Free parameters	32					24		
df	452					460		
Likelihood ratio	744					1165		
(Prob. perfect fit)	(<.01)					(<.01)		
RMSEA Index	.045					.069		
(Prob. close fit)	(<.92)					(<.01)		

Note. This table is based on grade-partialled data with N = 330 from Tables 2 and 3 and maximum-likelihood estimates from LISREL-8 (and Mx-92). An asterisk indicates parameter that is larger than 1.96 times its standard error; equal sign indicates a parameter has been fixed for model identification; less-than sign indicates a parameter that was restricted to a boundary. Basis $B[t]$ fixed equal to linear trend with 1-year proportion (i.e., t/12). Y_1 = Memory for Sentences; Y_2 = Memory for Words; Y_3 = Letter–Word Identification; Y_4 = Passage Comprehension. Prob = probability of; RMSEA = root mean square error of approximation.

roade, 1994). This final latent change result may be our most informative.

Discussion

Theoretical Issues

The psychometric evaluation of a test and the trait it measures is limited when made from only one occasion of measurement. A second time of measurement opens up some further possibilities but test–retest data are usually limited when the interval of time between tests is fixed at an arbitrary value. In these cases, test reliability is confounded with test-practice and other kinds of trait changes. In this article we used a varying time-lag test–retest interval to explore the separation of these components.

The time-lag design used here reinforces some well-known features of the differences between test reliability and trait stability. In the special case of parallel measures with equal means and equal variances, the internal consistency of a test can be estimated as the correlation between the two parallel measures. However, when these simplifying assumptions are not met (e.g., the observed variances over time are not equal) the factor-stability coefficient (R_{fs}) is not a substitute for, or counterpart of, the internal-consistency coefficient (R_{ic}). Similarly, if the trait scores change over time in a systematic way then the simple correlation over time $R_{y[1,2]}$ no longer reflects the same concepts about "continuity" over time (see McCall, Appelbaum & Hogarty, 1973; Wohlwill, 1973).

Two-occasion data provide the initial basis of the measurement of developmental change, even when additional measurements are obtained (Burr & Nesselroade, 1990). Choosing the most informative interval of time between these tests is a complex theoretical problem, which is not the same for all measures (see Gulliksen, 1950, p. 197; Cattell, 1957, pp. 343–344; Nunnally, 1978, p. 230). However, in theory, one reasonable use of the time-lag design will be at the beginning of an investigation when relatively little is known about the characteristics of the tests of the traits.

The main purpose of any structural equation analysis is to provide information about the validity of a theoretical construct (see McArdle & Prescott, 1992). As we have demonstrated here, tests measuring traits that change over time, whether as a result of the initial impact of practice or from longer term growth, will demonstrate lowered test–retest correlations. In these cases of systematic changes, it may be a serious mistake to assume that these lowered correlations are a reflection of lowered test quality. Of course, increases in growth are not the only possible explanation for a lowered test–retest correlation, and the identification of systematic growth remains an empirical issue.

Practical Issues

The practical implementation of a time-lag design can be relatively easy. Rather than measure the total sample at only one interval of time, the sample can be subdivided into different time-lag groups. Some previous research suggests that many different forms of incomplete data models can have reasonable power in these situations (McArdle & Hamagami, 1992). The resulting power to test basic growth hypotheses will vary as a function of the type of time-lag pattern selected, the number of occasions of measurement, and the communality of the variables used to indicate the common factors. Some researchers, most notably Schlesselman (1973) and Helms (1992), have pointed out both problems and benefits of time-lag designs.

In many studies, this time-lag data collection may serve to reduce the burden on the investigator. For example, not all participants need to be "retested in November" or "on each birthday." In other cases, this design may also mean that some increased burdens of data collection may now tend to fall on the investigator, especially if the design adds more sources of influence (i.e., confounds) than they were designed to rule out (i.e., control). In general, the practical utility of this time-lag design will vary among different kinds of psychological investigations (e.g., Bergmann, 1993; Cohen, 1991).

The time-lag design can provide some empirical basis for the determination of an optimal time-lag. In our illustrations, the relationships between the cognitive factors and other achievement cluster do vary over time, even with the relatively short daily and monthly time-lags. These results suggest some benefits in using longer time-lags between tests, especially for the cognitive factors. This time-lag approach might initially be used to determine a small enough aggregation of time-lag so we can pick up twice the hypothesized change patterns (i.e., the so-called "Nyquist limit"). When viewed as an empirical issue, the lowest level of aggregation may be desired and the model may best be fitted to individual level data (as in McArdle, 1994; McArdle & Hamagami, 1996). In many recent studies the specific time-lag between measurements is unplanned, unclear, or unreported. At very least, our highlighting of the time-lag may influence some researchers to consider these issues.

Final substantive results can also be practically presented in a form of the theoretical curves (see Figure 6) to illustrate the critical results of time-lag models. These parameters of any fitted model can also be used to form an expectation for the individuals involved in a specific testing. That is, we can calculate the likelihood of any specific vector of observed ($Y[t]_n$) scores compared with the expected group profiles. Of course, the individual inferences need to be made with appropriate caution. Any set of observed data, as presented in Figures 3 and 4 here, are not likely to be as simple or as smooth as the structural expectations of our theoretical models (e.g., Figure 6). In practice, it is likely that more elaborate models of developmental change will be needed to account for other important features of tests and traits.

Future Research Issues

It is also possible to consider more complex versions of this model where we estimate "an increasing growth function" and a "decreasing practice function" (see the Appendix). In related research (see McArdle & Hamagami, 1992, 1996), we have found that we can recover many additional parameters from these kinds of time-lag data, including multiple exponential and latent growth curve parameters. However, the power to detect differences among complex alternatives is greatly diminished when using only two occasions of measurement. More than two occasions, possibly using different time-lags, are indicated. In the WJ–R examples presented here, another retesting of some of these same individuals again at a later time would provide three time-points and this allows additional models not possible here (i.e., correlated components, more extended practice effects, more complex second-order coefficients, etc.).

Other aspects of these models can be expanded using nonconventional SEM. As noted previously, aggregated time-lag groups are not strictly needed because individual likelihood models can be written for individual time-lag distances (McArdle, 1994; Neale, 1993). When using a model with a more accurate account of the individual time between tests we found similar results. However, in general, an expanded time-lag data set should yield more informative and stable estimates (McArdle & Hamagami, 1992; Schlesselman, 1973). Further multivariate extensions allow the testing of many unique and informative hypotheses (e.g., Bergmann, 1993; Horn, 1972, 1988; McArdle, 1988; Woodcock, 1990).

The scientific utility of our time-lag extension of the test–retest design is an issue for future research. Any two-occasion test–retest design that incorporates a time-lag feature permits the structural separation of some potentially important developmental components. We think the WJ–R illustrations demonstrate that additional growth and change information can be substantively informative. The SEM approach presented here may be used with many other kinds of time-lag designs and data to help tease apart some interesting parameters related to reliability, stability, and change. We hope these ideas will be both useful in current practical applications and further extended in future theoretical developments.

References

Aber, M., & McArdle, J. J. (1991). Latent growth curve approaches to modeling the development of competence. In M. Chandler & M. Chapman (Eds.), *Criteria for competence: Controversies in conceptualization and assessment of children's abilities* (pp. 231–258) Hillsdale, NJ: Erlbaum.

Allison, P. (1987). Estimation of linear models with incomplete data. In C. Clogg (Ed.), *Sociological methodology, 1987* (pp. 71–103). San Francisco: Jossey-Bass.

Anastasi, A. (1954). *Psychological testing.* New York: McMillan.

Anderson, E. (1993). Analyzing change in short-term longitudinal research using cohort-sequential designs. *Journal of Consulting and Clinical Psychology, 61,* 929–940.

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24,* 145–152.

Bergmann, L. (1993). Some methodological issues in longitudinal research: Looking ahead. In G. Marcoulides & R. Schumacker (Eds.), *Longitudinal research on individual development: Present status and future perspectives* (pp. 217–241). Cambridge, England: University Press.

Blok, H., & Saris, W. (1983). Using longitudinal data to estimate reliability. *Applied Psychological Measurement, 7,* 295–301.

Bock, R., & Bargmann, R. (1966). Analysis of covariance structures. *Psychometrika, 31,* 507–533.

Browne, L., & Arminger, G. (1995). Specification and estimation of mean and covariance-structure models. In G. Arminger, C. Clogg, & M. Sober (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York: Plenum Press.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bolen & S. Long (Eds.), *Test-*

ing structural equation models (pp. 136–162). Newbury Park, CA: Sage.

Browne, M. W., & Du Toit, S. (1991). Models for learning data. In L. M. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 47–68). Washington, DC: American Psychological Association.

Bryk, A., & Raudenbush, S. (1993). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Burr, J., & Nessleroade, J. R. (1990). Change measurement. In A. Von Eye (Ed.), *Statistical methods in longitudinal research: Vol. 1. Principles and structuring change* (pp. 3–34). New York: Academic Press.

Cattell, R. (1957). *Personality and motivation structure and measurement.* Yonkers, NY: World Book.

Cattell, R. (1964). Validity and reliability: A proposed more basic set of concepts. *Journal of Educational Psychology, 55,* 1–22.

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

Cohen, P. (1991). A source of bias in longitudinal investigations of change. In L. M. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 18–25). Washington, DC: American Psychological Association.

Cronbach, L., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles.* New York: Wiley.

Gulliksen, J. (1950). *Theory of mental tests.* New York: Wiley.

Hartigan, J., & Wigdor, A. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Heise, D. (1969). Separating reliability and stability in test–retest correlation. *American Sociological Review,* 93–101.

Helms, R. (1992). Intentionally incomplete longitudinal designs: 1. Methodology and comparison of some full span designs. *Statistics in Medicine, 11,* 1189–1913.

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

Horn, J. L. (1972). State, trait, and change dimensions of intelligence. *The British Journal of Educational Psychology, 42,* 159–185.

Horn, J. L. (1988). Thinking about human abilities. In J. Nessleroade & R. B. Cattell (Eds.), *The handbook of multivariate experimental psychology* (Vol. 2, pp. 645–685). New York: Plenum Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGraw, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock–Johnson Technical manual: A reference on theory and current research* (pp. 197–246). Allen, TX: DLM Teaching Resources.

Horn, J. L., & Little, K. B. (1966). Isolating change and invariance in patterns of behavior. *Multivariate Behavioral Research, 1,* 219–229.

Horn, J. L., & McArdle, J. J. (1980). Perspectives on mathematical and statistical model building (MASMOB) in research on aging. In L. Poon (Ed.), *Aging in the 1980's: Psychological issues* (pp. 203–541). Washington, DC: American Psychological Association.

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

Hundal, P. S., & Horn, J. (1977). On the relationships between short-term learning and fluid and crystal intelligence. *Applied Psychological Measurement, 1,* 11–21.

Jones, M. (1962). Practice as a process of simplification. *Psychological Review, 69,* 274–294.

Jöreskog, K. G., & Sörbom, D. (1979). *Structural equation modeling.* Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 user's guide.* Chicago: Scientific Software.

Kiiveri, H. T. (1987). An incomplete data approach to the analysis of covariance structures. *Psychometrika, 52,* 539–554.

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

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

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

McArdle, J. J. (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). New York: Academic Press.

McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *The handbook of the psychology of aging* (Vol. 2, pp. 21–43). New York: Plenum Press.

McArdle, J. J., & Boker, S. M. (1990). *RAMpath: Path diagram software.* New York: Erlbaum.

McArdle, J. J., & Cattell, R. (1994). Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research, 29,* 63–113.

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

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

McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal data using latent growth structural models. *Experimental Aging Research, 18,* 145–166.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 89–124). Hillsdale, NJ: Erlbaum.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *The British Journal of Mathematical and Statistical Psychology, 37,* 234–251.

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. 223–267). Hillsdale, NJ: Erlbaum.

McArdle, J. J., & Prescott, C. (1992). Age-based construct validation using structural equation models. *Experimental Aging Research, 18,* 87–115.

McCall, R. B., Appelbaum, M.I., & Hogarty, P. S. (1973). Developmental changes in mental performance. *Monographs of the Society for Research in Child Development, 38,* (3, Serial No. 150).

McDonald, R. P. (1980). A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *British Journal of Mathematical and Statistical Psychology, 33,* 161–183.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

McGrew, K. S., Werder, J., & Woodcock, R. W. (1991). *WJ–R technical manual.* Chicago: Riverside Publishing.

Meredith, W. (1991). Latent variable models for studying differences and change. In L. M. Collins & Horn J. (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 149–163). Washington, DC: American Psychological Association.

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

Muthen, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431–462.

Neale, M. C. (1993). *Mx statistical modeling.* Unpublished computer program manual, Virginia Commonwealth University, Department of Human Genetics, Medical College of Virginia, Richmond.

Nesselroade, J. R. (1983). Temporal selection and factor invariance in the study of development and change. In J. R. Nesselroade & P. Baltes (Eds.), *Life-span development and behavior* (Vol. 5, pp. 59–87). New York: Academic Press.

Nesselroade, J. R., & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development.* New York: Academic Press.

Nesselroade, J. R., & Bartsch, T. (1977). Multivariate perspectives on the construct validity of the trait–state distinction. In R. Cattell & R. Dreger (Eds.), *Handbook of modern personality theory* (pp. 221–238). Washington, DC: Hemisphere Press.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Raffalovich, L., & Bohrnstedt, G. (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data. *Sociological Methods and Research, 15,* 385–405.

Raykov, T. (1993). A structural equation model for measuring change and discerning patterns of growth and decline. *Applied Psychological Measurement, 17,* 53–71.

Reise, S., Widaman, K., & Pugh, R. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92,* 726–748.

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

Saris, W., & Satorra, A. (1993). Power evaluations in structural equation models. In K. Bolen & S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Thousand Oaks, CA: Sage.

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64,* 92–107.

Schlesselman, J. J. (1973). Planning a longitudinal study: 2.

Frequency of measurement and study duration. *Journal of Chronic Disease, 26,* 561–570.

Shavelson, R., & Webb, N. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 356–442). Washington, DC: American Council on Education.

Steyer, R. (1989). Models of classical psychometric test theory as stochastic models: Representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika, 3,* 25–60.

Steyer, R., Schwenkmezger, P., & Auer, A. (1990). The emotional and cognitive components of trait anxiety: A latent state–trait model. *Personality and Individual Differences, 11,* 125–134.

Thissen, D., & Steinberg, L. (1988). Data analysis using items response theory. *Psychological Bulletin, 104,* 385–395.

Thorndike, R. (1931). The effect of the interval between test and retest on the constancy of the IQ. *Journal of Educational Psychology,* 1–7.

Traub, R. (1994). *Reliability for the social sciences: Theory and applications.* Newbury Park, CA: Sage.

Tucker, L. R., Damarin, F., & Messick, S. (1966). A base-free measure of change. *Psychometrika, 31,* 457–473.

Vinsonhaler, J., & Meredith, W. (1966). A stochastic model for repeated testing. *Multivariate Behavioral Research, 4,* 461–477.

Werts, C. E., Breland, H. M., Grandy, J., & Rock, D. (1980). Using longitudinal data to estimate reliability in the presence of correlated measurement errors. *Educational and Psychological Measurement, 40,* 19–29.

Whitman, W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (Vol. 2, pp. 505–560). New York: Plenum Press.

Willett, J. (1988). Questions and answers in the measurement of change. *Review of Research in Education, 15,* 345–422.

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

Woodcock, R. W. (1990). Theoretical foundations of the WJ–R measures of cognitive ability. *Journal of Psycho-Educational Assessment, 8,* 231–258.

Woodcock, R. W. (1993). An information processing view of Gf-Gc theory. *Journal of Psycho-Educational Assessment,* 80–101.

Woodcock, R. W., & Johnson, M. B. (1989). *The Woodcock–Johnson Psycho-Educational Battery Revised.* Chicago: Riverside.

Woodworth, R. (1938). *Experimental psychology.* New York: Holt.

Wright, S. (1982). On "Path analysis in genetic epidemiology: A critique." *American Journal of Human Genetics, 35,* 757–762.

Appendix
Technical Notes

Notes on the Woodcock–Johnson—Revised (WJ–R) Data

The WJ–R scales are a wide-range comprehensive set of individually administered tests of intellectual ability, scholastic aptitudes, and achievement (McGrew et al., 1991; Woodcock, 1990). Four features of the WJ–R make it especially valuable as an instrument for research in human development and psychometric change: The WJ–R (a) is well-normed, (b) is calibrated using a Rasch model, (c) includes multiple ability measures, and (d) can be administered quickly and easily. All WJ–R scales use a constant of 500 and a logit transformation so a change of 10 points indicates a 25% difference in probability of correct response. The equal interval feature of these Rasch-based scales is useful in time-lag research. In theory, test differences can be interpreted to have the same meaning at any performance level.

In Study 1 presented here, all participants were initially administered the WJ–R Memory for Names task. This task is basically designed to measure the long-term memory of the participant. In the first testing, participants were asked to examine a set of figures called space creatures and to memorize their fictitious names (i.e., "This is Jawf. Point to Jawf. This is Kiptron. Point to Kiptron," etc.). Progressively more figurines are added to the pictures and the task becomes more difficult. Raw scores on the first testing session reflect the maximum number (0 to 12) of names held in memory at any time during the session. The retest component of this experiment occurred sometime between 1 and 14 days later, with an average lag of about 3 days. In all cases the same participant was again asked to name the same space creatures. This task was again designed to measure the long-term memory of the participant. Three attempts were allowed to name each space creature (so raw scores range between 0 and 36). Both test and retest scores obtained were converted to a Rasch-based measurement scale: In these units the average raw score was 500.5 at the first occasion, 490.5 at the second occasion, and the overall test–retest correlation was $R_{y[1,2]} = .898$.

In Study 2 we selected a stratified random sample of individuals from the same norming sample for a longer term retesting. Out of 402 students contacted, 361 (89.9%) agreed to be tested again (245 kindergarten–Grade 12 and 116 college students) and 330 students had complete data. This sampling approach resulted in an average retest delay between tests of 245 days, with a minimum of 21 days, a maximum of 482 days, and a small correlation between age and time-lag ($R_{a,t} = -.14$).

All WJ–R scores used here were age-adjusted residuals from a fourth-order polynomial model:

$$Y[t]_n = B_0 + B_1 \times X_n + B_2 \times X_n^2 + B_3 \times X_n^3 + B_4 \times X_n^4 + \epsilon_n, \qquad \text{(A1)}$$

where X = age in years − 10, and where coefficients B are taken from the larger WJ–R norming sample and applied to each score at each time point. It follows that the intercepts $M_{i,w}$ are artifacts of the age-adjusted equations (and these are not included in the diagrams). This adjustment was applied so we would not overestimate the test–retest correlation due to persons remaining about to the same age during these experimental treatments. The numerical results show that score changes are largely linear over this age span, so this age adjustment reduced the variance at the initial time-point and the test–retest correlation for all groups. There was no other substantial difference in the models fit before and after this age adjustment. Although there may still be Age × Retest interactions, this simple adjustment procedure allows us to focus on test–retest effects here.

In some of the models presented here, an estimate of the unique variance was used as a fixed value as an approximation of the "disattenuated" correlation. That is, the fixed unique or error variance estimate was calculated from $D_u = D[1] \times \sqrt{1 - R_{ic}}$ estimated from the larger sample. McGrew et al. (1991) reported internal consistency reliabilities in the norming study (median $R_{ic} = .947$ and $R_{ic} = .888$), and we use these as initial estimates of the loadings and uniquenesses here. In the multivariate models we use the corresponding summary statistics for the individual scales in each composite (i.e., MEMSEN, MEMWRD, LWIDNT, and PSGCMP)—these statistics were not all listed in Tables 2 and 3. In each of these models the factor loadings and uniquenesses was estimated from the time-lag data.

Latent Means, Covariances, and Common Factor Notation

To define the latent means and covariances we write

$$\begin{aligned}
&\mathscr{E}\{I\} = M_i, \mathscr{E}\{(I - M_i)(I - M_i)'\} = V_i \\
&\mathscr{E}\{G\} = M_g, \mathscr{E}\{(G - M_g)(G - M_g)'\} = V_g, \\
&\mathscr{E}\{P\} = M_p, \mathscr{E}\{(P - M_p)(P - M_p)'\} = V_p, \\
&\mathscr{E}\{U\} = 0, \mathscr{E}\{U\,U'\} = V_u,
\end{aligned}$$

and

$$\mathscr{E}\{(I - M_i)(G - M_g)'\} = C_{ig}. \qquad \text{(A2)}$$

The factor-analytic basis of the latent growth model has been discussed in other research (e.g., see McArdle, 1988; Meredith & Tisak, 1990; Browne & Arminger, 1995). We can expand Model 3 for a specific time-series data (e.g., $t = 0$ to 4) as

$$\begin{aligned}
Y[0]_n &= I_n + B[0] \times G_n + A[0] \times P_n + U[0]_n, \\
Y[1]_n &= I_n + B[1] \times G_n + A[1] \times P_n + U[1]_n, \\
Y[2]_n &= I_n + B[2] \times G_n + A[2] \times P_n + U[2]_n, \\
Y[3]_n &= I_n + B[3] \times G_n + A[3] \times P_n + U[3]_n, \\
Y[4]_n &= I_n + B[4] \times G_n + A[4] \times P_n + U[4]_n. \qquad \text{(A3)}
\end{aligned}$$

These vectors can now be summarized into the more compact matrix form described in Equations 4 and 5. In this specific factor model with $T = 5$ we include a (5 × 3) matrix $\mathbf{L}$ of common factor loadings, a (3 × 1) vector of common factor scores $\mathbf{Q} = [\mathbf{I_n}, \mathbf{G_n}, \mathbf{P_n}]$, and a (5 × 1) vector of independent unique scores $\mathbf{U} = \mathbf{U[t]}$ (for $t = 0$ to 4).

In the factor-analytic context we can also write a matrix form of the latent means and covariances. An average cross-products or moment matrix among the common factors $\mathbf{Q}$ can be written as

$$\mathbf{M_{qq}} = \begin{bmatrix} M_i^2 + V_i & & sym. \\ M_g \times M_i + C_{ig} & M_g^2 + V_g & \\ M_p \times M_i & M_p \times M_g & M_p^2 + V_p \end{bmatrix}, \qquad \text{(A4)}$$

and we define the moment matrix of unique variances as

$$\mathbf{M_{uu}} = \begin{bmatrix} V_u & & & & sym. \\ & V_u & & & \\ & & V_u & & \\ & & & V_u & \\ & & & & V_u \end{bmatrix}. \qquad \text{(A5)}$$

So, by assuming the independence of the $\mathbf{Q}$ and $\mathbf{U}$ we can write the usual structural factor analysis representation

$$\mathbf{M_{yy}} = \mathbf{L} \times \mathbf{M_{qq}} \times \mathbf{L}' + \mathbf{M_{uu}}, \qquad \text{(A6)}$$

and create expectations about the mean and covariances (moments) of all observed variables.

It is also possible to add the defined unit constant variable into the vector of latent variables. This has the advantage of directly separating the means from the moments, and this covariance-based form produces identical structural expectations $\mathbf{M_{yy}}$. This approach also permits estimation by standard SEM programs based on covariance matrices (for references, see McArdle, 1988).

Multiple Group Estimation With Incomplete Data

The expected covariance matrix and mean vectors for W observed variables and K latent variables can always be formed using RAM formulas (see McArdle & McDonald, 1984; McDonald, 1985) as,

$$\Sigma = \mathbf{F}(\mathbf{I}-\mathbf{A})^{-1}\,\mathbf{S}(\mathbf{I}-\mathbf{A})^{-1\prime}\,\mathbf{F}' \text{ and}$$
$$\mu = \mathbf{F}(\mathbf{I}-\mathbf{A})^{-1}\,\mathbf{J}, \qquad \text{(A7)}$$

where Σ = the ($W \times M$) expected covariance matrix, μ = the ($W \times 1$) expected mean vector, $\mathbf{F}$ = the ($W \times [W + K]$) filter matrix, $\mathbf{A}$ = the (($W + K$) $\times$ ($W + K$)) assymmetric coefficient matrix, $\mathbf{S}$ = the ([$W + K$] $\times$ [$W + K$]) symmetric covariance matrix, and $\mathbf{J}$ = the ([$W + K$] $\times$ 1) latent mean coefficient matrix. Some uses of these matrices are described below.

For any set of parameter values the likelihood ratio test of the difference between the expected and observed covariance matrix and mean vectors can be formed by calculating

$$F_{expected} = -2LL_{\Sigma} = -k * ln(2\pi) - ln\,|\Sigma| + \mu'\Sigma^{-1}\,\mu,$$

and

$$F_{observed} = -LL_C = -k * ln(2\pi) - ln\,|\mathbf{C}| + \mathbf{M}'\mathbf{C}^{-1}\,\mathbf{M},$$

so

$$\begin{aligned} LRT &= (N-1)\,[F_{observed} - F_{expected}] \\ &= ln\,|\Sigma| + ln\,|\mathbf{C}| + tr\,(\Sigma^{-1}\,\mathbf{C}) - p \\ &\quad + (\mathbf{M}-\mu)'\Sigma^{-1}\,(\mathbf{M}-\mu) \approx \chi^2_{df(d)}. \end{aligned} \qquad \text{(A8)}$$

In the case of multiple independent groups, the usual likelihood function is weighted by the appropriate sample size by calculating

$$LRT = \sum_{g=1}^{G} [(N-1)^{(g)}\,(F_{observed}{}^{(g)} - F_{expected}{}^{(g)})] \approx \chi^2_{df(d)} \qquad \text{(A9)}$$

Several other tests of goodness-of-fit (e.g., Browne & Cudeck, 1993) statistical power analyses (e.g., McArdle, 1994) can now be formed from these indices.

SEM Programming Devices

Various computer programs used in this article can all be obtained as ASCII files under the title of JJM.TIMELAG96 from the Anonymous FTP server at the University of Virginia (FTP FTP.VIRGINIA.EDU). These formal models can be analyzed by both the LISREL-8 computer program (Joreskog & Sorbom, 1993) and the Mx program (Neale, 1993). Both programs allow us to write patterns for the means, deviations, and correlations of the time-lag groups (as described in Tables 1, 2, and 3). These patterns are defined by parameters that are (a) free to be estimated using numerical procedures, (b) fixed at some specific value, or (c) equal to to another parameter (i.e., invariant).

A slightly nonstandard matrix approach to model representation was used to simplify all models here. The univariate latent growth model of Figure 5 was fitted to eight groups (of Table 1) in the following way. First, a matrix specification was set up for each group with three observed variables ($Y[0]$, $Y[t]$, and one unit constant), and 14 total variables (all variables in Figure 5). We used a fixed-filter matrix $\mathbf{F}$ (3 × 14) containing only ones and zeros, an asymmetric regression matrix $\mathbf{A}$ (14 × 14) containing all one-headed arrows, and a symmetric covariance matrix $\mathbf{S}$ (14 × 14) containing all two-headed arrows. With this approach the matrix entries precisely match the parameters in the path diagram of Figure 5 and, although the program is a bit slow, it is easy to use the input and output.

This same model was then used to specify the one-factor model simply by placing zeros in the appropriate rows and columns. In multivariate models, parameters representing the factor mean vector ($\mathbf{J}$) are needed.

This kind of incomplete data take advantage of the special use of the RAM filter matrix. Then for each group, we write a separate filter matrix that defines the available measurements for that specific group. For example, if we have a unit constant and two measurements at t = 0 and t = 1 and only eight total variables, we would write

$$\mathbf{F}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \text{(A10)}$$

but in a group with a unit constant and two measurements at t = 0 and t = 2, we write

$$\mathbf{F}^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \text{(A11)}$$

and in a group with a unit constant and two measurements at t = 0 and t = 7, we write

$$\mathbf{F}^{(7)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \qquad \text{(A12)}$$

In general, the placement of the unit value in the last row indicates the available data for each time-lag group, and this is the only parameter that is altered from one group to another. Following McArdle and Anderson (1990) we write one set of model parameters in super-matrices $\mathbf{A}$ and $\mathbf{S}$. This approach allows the specification of invariance of all other model parameters in matrices $\mathbf{A}^{(g)}$ and $\mathbf{S}^{(g)}$. There are numerous ways to produce the correct model expectations, but these require much more complex programming.

The calculation of standard errors for the variance components poses a special problem that was not detailed in the previous sections. In general, the variance terms (e.g., V_i) have an asymmetric distribution, so we estimated their confidence intervals by estimating the comparable standard deviations (e.g., D_i) and their standard errors. The calculation of confidence intervals for the standardized variance components (e.g., V_i^*) is more complex. These proportions include several model parameters and the correlations among these estimates need to be taken into account as well (as in the calculation of the standard errors for indirect effects). These standard error calculations can be built into the model estimation by adding extra parameters to the models (i.e., the PAR command) and then using nonlinear constraints (i.e., the CO commands) to form these ratios.

Time-Lag Model Mathematical Expectations

We examined all expectations plotted in Figure 6 using the matrix expressions defined above. For example, if we substitute the parameters of loading matrix $\mathbf{L}_3$ we can write the observed means as

$$\begin{aligned}\mathcal{E}\{M_{y[0]}\} &= M_i + 0 \times M_g + 0 \times M_p,\\ \mathcal{E}\{M_{y[1]}\} &= M_i + 1 \times M_g + 1 \times M_p,\\ \mathcal{E}\{M_{y[2]}\} &= M_i + 2 \times M_g + 1 \times M_p,\\ \mathcal{E}\{M_{y[3]}\} &= M_i + 3 \times M_g + 1 \times M_p,\\ \mathcal{E}\{M_{y[4]}\} &= M_i + 4 \times M_g + 1 \times M_p, \end{aligned} \tag{A13}$$

which illustrate the functional relationships over time. If we further define $M_i = 1$, $M_g = 1$, and $M_p = 1$, then, by simple substitution, these expectations yield means $M_{y[0]} = 1$, $M_{y[1]} = 3$, $M_{y[2]} = 4$, $M_{y[3]} = 5$, $M_{y[4]} = 6$, and these are the numerical values plotted for $\mathcal{E}_3$ in Figure [6a].

All other variances, correlations, and variance proportions listed in Figure 6 were created by substitution in the same way. The parameters used to create these four models are listed in the design outlined in Table 1A:

Multivariate Time-Lag Expectations

The multivariate path diagram may be written algebraically in a number of ways. To simplify matters here, we have written the multivariate expectations needed as separate elements.

Structural expectations for the means may be written for variable w as

$$\mathcal{E}\{M_{yw[0]}\} = H_w \times M_{f[0]} + M_{ew},$$

and

$$\mathcal{E}\{M_{yw[t]}\} = H_w \times M_{f[t]} + A[t]_w M_{pw} + M_{ew}$$

where

$$M_{f[t]} = M_i + B[t] \times M_g. \tag{A14}$$

Structural expectations within variables (for measure w) may be written as

$$\begin{aligned}\mathcal{E}\{V_{yw[0]}\} &= H_w^2 \times V_{f[0]} + V_{ew},\\ \mathcal{E}\{V_{yw[t]}\} &= H_w^2 \times V_{f[t]} + A[t]_w^2 \times V_{pw} + V_{ew},\end{aligned}$$

and

$$\mathcal{E}\{C_{w[0,t]}\} = H_w^2 \times C_{f[0,t]}$$

where

$$V_{f[t]} = V_i + B[t]^2 \times V_g + 2\, B[t] \times C_{ig} + V_s$$

and

$$C_{f[0,t]} = V_i + B[t] \times C_{ig}. \tag{A15}$$

Structural expectations among variables (for measures $j \neq k$) may be written as

$$\begin{aligned}\mathcal{E}\{C_{j[0],k[0]}\} &= H_j \times (V_i + V_s) \times H_k',\\ \mathcal{E}\{C_{j[t],k[t]}\} &= H_j \times (V_i + B[t]^2 \times V_g + 2\, B[t]\\ &\quad \times C_{ig} + V_s) \times H_k'\end{aligned}$$

and

$$\mathcal{E}\{C_{j[0],k[t]}\} = H_j \times (V_i + B[t] \times C_{ig})\, H_k'. \tag{A16}$$

A few test-specific developmental ratios within the factor model may be written as

$$R_{fc(yw[t])} = \frac{H_w^2 \times V_{f[t]}}{V_{yw[t]}}$$

and

$$R_{pr(yw[t])} = \frac{A[t]_w^2 \times V_p w}{V_{yw[t]}}. \tag{A17}$$

A few trait-specific developmental ratios for the factor model may be written as

Table 1A
The Numerical Values for the Four Theoretical Models Presented in Figure 6

	Loadings			Means			Variances				Correlations
Model label	H	B_t	A_t	M_i	M_g	M_p	V_i	V_g	V_p	V_u	C_{ig}
$\mathcal{E}_0$	1	0	0	1	0	0	3	0	0	2	0
$\mathcal{E}_1$	1	t/12	0	1	1	0	3	2	0	2	0
$\mathcal{E}_2$	1	0	$e^{-.2(t-1)}$	1	0	1	3	0	1	2	0
$\mathcal{E}_3$	1	t/12	1	1	1	1	3	2	1	2	0

$$R_{fs[t]} = \frac{V_i}{V_{f[t]}}, \tag{A18}$$

$$R_{fg[t]} = \frac{B[t]^2 \times V_g \times B[t] \times C_{ig}}{V_{f[t]}},$$

and

$$R_{sf[t]} = \frac{V_s}{V_{f[t]}}.$$

More Advanced Growth Functions

Parameter identification in this kind of a model can be achieved using various devices. In a model with k common factors we know that at least k^2 constraints need to be fixed and distributed across the loadings and covariances of the common factors (for details, see McArdle & Cattell, 1994, among others).

In the specific growth models discussed above, two prior conditions are clear. First, the initial factor I has a fixed scale of measurement simply by the definition of the fixed unit values in the first column. Second, for the first factor to be interpreted as the initial level we also need to set the other loadings for the initial time point to zero (i.e., $L_{1,2} = B[0] = 0$ and $L_{1,3} = A[0] = 0$). The necessary scaling of each of the other two factors can conveniently be defined by restrictions where both $B[1] = 1$ and $A[1] = 1$. This leaves an initially restricted model where the second and third columns can be estimated for $t > 1$.

The main problem with this model is that second and third common factors, G and P, are not yet separated. To wit, if $B[2] = A[2]$, $B[3] = A[3]$, and $B[4] = A[4]$, then the columns are identical and could be interchanged without loss of meaning (the matrix rank is greater than the matrix order). This means the remaining loadings require additional restrictions or the overall model will not be identified. Following a suggestion made by McDonald (1980), we might consider a representation of this model where we estimate "an increasing growth function" and a "decreasing practice function" by writing

$$L_m = \begin{bmatrix} 1 & 0 & 0 \\ 1 & B[1]=1 & A[1]=1 \\ 1 & B[2]=B[1]+\delta[1]^2 & A[2]=A[1]-\theta[1]^2 \\ 1 & B[3]=B[2]+\delta[2]^2 & A[3]=A[2]-\theta[2]^2 \\ 1 & B[4]=B[3]+\delta[3]^2 & A[4]=A[3]-\theta[3]^2 \end{bmatrix}. \tag{A19}$$

In this model each growth coefficient $B[t] > B[t-1]$ due to the positive increment $\delta[t-1]^2$, so the growth function must be monotonically increasing or flat. Likewise, each practice coefficient $A[t] < A[t-1]$ due to the negative increment $-\theta[t-1]^2$, so the practice function must be monotonically decreasing or flat. In this way, we have a model where one function describes the increases over time and another function describes the decreases over time.

Received May 16, 1996
Revision received March 29, 1997
Accepted April 9, 1997 ■