Developmental Psychology
2002, Vol. 38, No. 1, 115–142

0012-1649/02/$5.00 DOI: 10.1037//0012-1649.38.1.115

Comparative Longitudinal Structural Analyses of the Growth and Decline of Multiple Intellectual Abilities Over the Life Span

John J. McArdle, Emilio Ferrer-Caja, Fumiaki Hamagami, and Richard W. Woodcock
University of Virginia

Latent growth curve techniques and longitudinal data are used to examine predictions from the theory of fluid and crystallized intelligence (*Gf-Gc* theory; J. L. Horn & R. B. Cattell, 1966, 1967). The data examined are from a sample ($N \sim 1{,}200$) measured on the Woodcock-Johnson Psycho-Educational Battery—Revised (WJ–R). The longitudinal structural equation models used are based on latent growth models of age using two-occasion "accelerated" data (e.g., J. J. McArdle & R. Q. Bell, 2000; J. J. McArdle & R. W. Woodcock, 1997). Nonlinear mixed-effects growth models based on a dual exponential rate yield a reasonable fit to all life span cognitive data. These results suggest that most broad cognitive functions fit a generalized curve that rises and falls. Novel multilevel models directly comparing growth curves show that broad fluid reasoning (*Gf*) and acculturated crystallized knowledge (*Gc*) have different growth patterns. In all comparisons, any model of cognitive age changes with only a single *g* factor yields an overly simplistic view of growth and change over age.

A great deal of prior substantive research on cognitive abilities has provided information about the growth and decline of intellectual abilities with data collected over the full life span. In this study, we present theory about the growth and change over age in different cognitive abilities (e.g., Cattell, 1941, 1998; Horn, 1988, 1998; Swanson, 1999). We examine these theories using contemporary statistical models of a few classical questions: How do broad cognitive functions grow and change within an individual over age and time? and Are these growth and change patterns different from one variable to another? We fit several new models to two-occasion longitudinal data from a relatively large sample of persons ($N \sim 1{,}200$) measured on multiple cognitive tests from the Woodcock-Johnson Psycho-Educational Battery—Revised (WJ–R; McGrew, Werder, & Woodcock, 1991; Woodcock & Johnson, 1989).

A great deal of prior methodological research has focused on mixing cross-sectional and longitudinal data. In recent research, McArdle and Woodcock (1997) used latent growth structural equation models to examine cognitive test–retest data collected over varying intervals of time—that is, a time-lag design. Because age variation was not the focus of that study, age scores were partialed (using fourth-order polynomials) before the time-lag model was fitted, and linear and nonlinear practice components were used. These time-lag analyses showed that these intellectual ability traits had high factor stability, that some traits exhibited important trait changes, that all tests used had very small practice components, and that no single general factor was evident in the patterns over time.

In the current research, we extend these previous latent growth models to describe the potential benefits of a mixture of age-based and time-based models that use only two time points of data collection—that is, an accelerated longitudinal design (Aber & McArdle, 1991; Bell, 1954; McArdle & Anderson, 1990; McArdle & Bell, 2000; McArdle & Woodcock, 1997)—especially for studies across the life span (e.g., S. C. Duncan & Duncan, 1995; McArdle & Hamagami, 1992; cf. Swanson, 1999). This article presents some practical ways to fit latent growth models of age with incomplete data, using standard multilevel software (e.g., SAS PROC MIXED). We also expand these models to include nonlinear growth patterns and comparisons of growth patterns of difference (e.g., SAS PROC NLMIXED).

Substantive Predictions From the Theory of Fluid and Crystallized Intelligence

The theory of fluid and crystallized intelligence (i.e., *Gf-Gc* theory; Cattell, 1941, 1971, 1987; Horn, 1971, 1988, 1998; Horn & Cattell, 1966, 1967) proposes that primary abilities are structured into two principal dimensions, namely, fluid (*Gf*) and crystallized (*Gc*) intelligence. The first common factor, *Gf*, represents a measurable outcome of the influence of biological factors on intellectual development (i.e., heredity, injury to the central nervous system), whereas the second common factor, *Gc*, is considered the main manifestation of influence from education, experience, and acculturation. *Gf-Gc* theory disputes the notion of a unitary structure, or general intelligence, as well as, especially in the origins of

---

John J. McArdle, Emilio Ferrer-Caja, Fumiaki Hamagami, and Richard W. Woodcock, Department of Psychology, University of Virginia.

This research was supported by grants from the National Institute on Aging (AG02695, AG04704, and AG07137). This work would not have been possible without the Measurement/Learning/Consultants personnel who carried out the experimental design, especially Mary Reuf, and the cooperation of many other National Growth and Change Study testers and subjects. We also thank our colleagues John L. Horn, John R. Nesselroade, Karen M. Schmidt, and Carol A. Prescott for their assistance in this research and for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to John J. McArdle, Department of Psychology, P.O. Box 400400, University of Virginia, Charlottesville, Virginia 22904-4400. E-mail: jjm@virginia.edu

the theory, the idea of a structure comprising many restricted, slightly different abilities. From these substantive premises, *Gf-Gc* theory makes a few explicit predictions about the complex nature of human intellectual abilities in three related areas:

1. The first predictions are *structural*: A single general factor (i.e., *g* from Spearman, 1904) will not account for the patterns of variation seen among multiple abilities—at least two broad factors are required for a reasonable level of fit to observations. One broad factor, *Gc*, was thought to represent acculturated knowledge, and the other broad factor, *Gf*, was thought to represent reasoning and thinking in novel situations.

2. The second set of predictions of *Gf-Gc* theory concerns *kinematic* trends: Over the early phases of the life span, there is an expected rise of *Gc* together with an expected rise of *Gf*, but in early adulthood, there is further growth of *Gc* while the *Gf* peaks early and rapidly declines in older ages.

3. A third set of predictions is about *dynamic* processes and interrelationships among factors: There is an "investment" of *Gf* coupled with other lower order factors in the context of educationally relevant settings, and this investment leads to individual differences in the development of *Gc*.

These general predictions have been examined by different researchers using many different experimental methods (see Anstey, Luszcz, & Sanchez, 2001; Carroll, 1993, 1998; Flanagan & McGrew, 1996; Horn, 1988, Horn, 1998; McArdle, Hamagami, Meredith, & Bradway, 2000; McArdle & Prescott, 1992; McGrew & Flanagan, 1998; Woodcock, 1990). For example, regarding the structural predictions, Carroll (1993) argued that the fluid–crystallized structure is only a second stratum in a more complex three-stratum structure. At the base of this hierarchy, or at the first stratum, Carroll located many primary abilities, whereas he considered the third stratum a general factor, *g*, the result of the common factor variance of the second-stratum factors. Of key interest in this study are the age-curve predictions of *Gf-Gc* theory, and these have generally found support in a wide range of recent cross-sectional studies (e.g., Horn & Cattell, 1967; Horn & Noll, 1997; Lindenberger & Baltes, 1997; McArdle & Prescott, 1992). The prominent feature of a developmental separation of *Gf* and *Gc* factors is depicted here in Figure 1 (from Cattell, 1987). This theoretical plot has both *Gf* and *Gc* functions rising through youth until early adulthood, when *Gf* declines most rapidly while *Gc* is continuously rising well into the 60s and 70s. Also illustrated here is the classical problem of creating a composite of the two, termed *traditional intelligence*. Key developmental information is lost because of aggregation (i.e., averaging) over the two independent constructs.

Empirical evidence in support of this broad separation of cognitive functions over the life span predated the initial statements of *Gf-Gc* theory (Cattell, 1971, pp. 186). In particular, researchers in this area recognized Jones and Conrad's (1933) seminal research, a well-known and large-scale cross-sectional study of the Army Alpha tests. Jones and Conrad (1933) concluded,

> The chief characteristic of the curve may be summarized as involving a linear growth to about 16 years, and a negative acceleration beyond 16 to a peak between the ages of 18 and 21. A decline follows, which is much more gradual than the curve of growth, but which by the age of 55 involves a recession to the 14-year level. (p. 239)

Jones and Conrad (1933) also recognized another key result:

> Of special interest is the observation that the tests showing the most rapid decline are Tests 7 (analogies), 3 ("common sense"), and 6 (numerical completion). These tests may perhaps be considered, at least on *a priori* grounds, to be the best in the Army Alpha for the measurement of basic intelligence, i.e., to be most free from the influence of environmental variables, and from the accumulative effects of differential experience. Our results here confirm Thorndike's conclusion that age exerts its most adverse influence upon native capacity or "sheer modifiability." (p. 253)

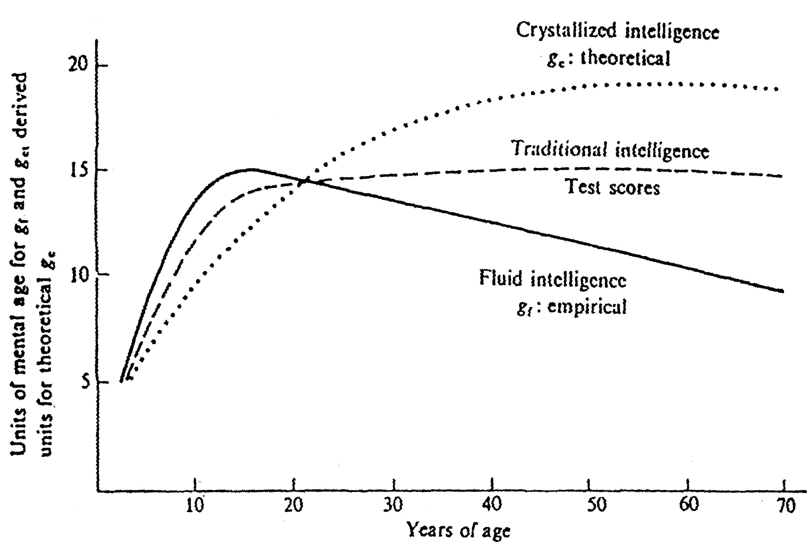

Figure 1. A theoretical description of life span curves of intellectual abilities. From *Intelligence: Its structure, growth and action* (p. 206) by R. B. Cattell, 1987, Amsterdam: North-Holland. Copyright 1987 by Elsevier Science Publishers. Reprinted with permission.

In a summary of prior research, Botwinick (1977) referred to this distinction in the growth patterns of different intellectual abilities as the "classic aging pattern," and others labeled it as the "hold versus no-hold pattern" (e.g., Bayley, 1966; Eichorn, Clausen, Haan, Honzik, & Mussen, 1981; Horn, 1970; W. L. Hunt, 1949; Wechsler, 1955). The empirical validity of the classic aging pattern has been brought into question by the results of studies in which abilities measured with the "hold" tests were found to decline with age as well (e.g., Baltes & Schaie, 1976; Kangas & Bradway, 1971; Wechsler, 1955). The separate age-curve predictions of *Gf-Gc* theory have also been examined in detail in longitudinal studies, and these have provided more recent empirical support (e.g., Baltes & Mayer, 1999; Donaldson & Horn, 1992; McArdle, Prescott, Hamagami, & Horn, 1998; Schaie, 1996).

The measurement basis of *Gf-Gc* theory has been expanded considerably in work by Horn (1970, 1972, 1988, 1991, 1998). Much of this work attempted to bring together relevant results from experimental studies of cognition with results from studies of individual differences. A path diagram summary of the broad structural factors in the contemporary version of this cognitive systems theory is presented here in Figure 2 and includes 8 to 10 common factors (for details, see Horn, 1985), each measurable using multiple indicators. Some key broad cognitive abilities within this system can be defined as follows:

Fluid reasoning (Gf)—the ability to reason, form concepts, and solve problems that often involve unfamiliar information or procedures. *Gf* is manifested in the reorganization, transformation, and extrapolation of information.

Comprehension–knowledge (Gc)—the breadth and depth of knowledge, including verbal communication and information. When previously learned procedures are used, reasoning is also included.

Long-term retrieval (Glr)—the ability to store information efficiently and retrieve it later through association.

Short-term memory (Gsm)—the ability to hold information in immediate awareness and then use it within a few seconds; also related to working memory.

Visual processing (Gv)—spatial orientation, the ability to analyze and synthesize visual stimuli, and the ability to hold and manipulate mental images.

Auditory processing (Ga)—the ability to discriminate, analyze, and synthesize auditory stimuli. *Ga* is also related to phonological awareness.

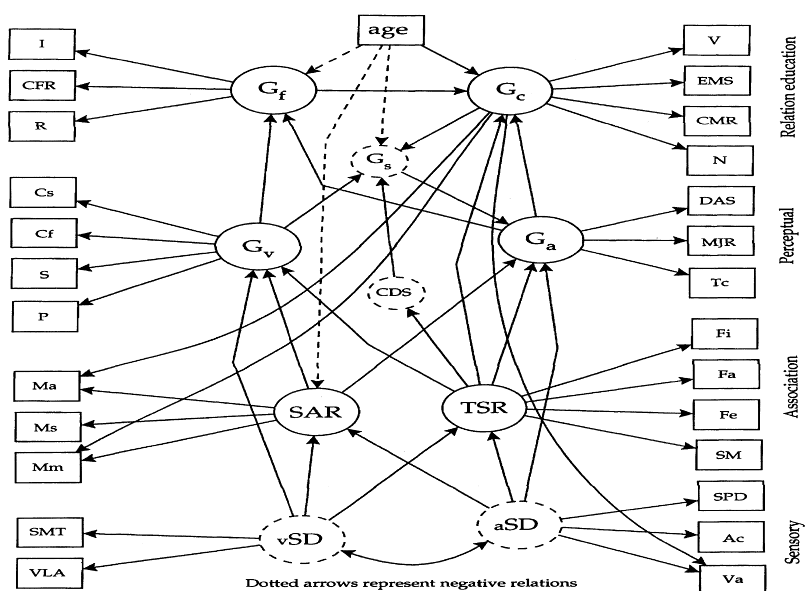

Figure 2. A multiple-factor depiction of contemporary *Gf-Gc* theory. G_f = Fluid Ability, I = inductive reasoning, CFR = cognition of figural relations, R = quantitative reasoning; G_v = Broad Visualization, Cs = speed of closure, Cf = flexibility of closure, S = spatial orientation, P = perceptual speed; SAR = Short-Term Acquisition Retrieval, Ma = associative memory, Ms = span immediate memory, Mm = meaning paired associates immediate memory; vSD = Visual Sensory Detectors, SMT = sperling matrix awareness, VLA = visual location address; G_c = Crystallized Ability, V = verbal comprehension, EMS = evaluation of semantic systems, CMR = cognition of semantic relations, N = number facility; G_a = Broad Auditory Thinking, DAS = discriminate patterns of sounds, MJR = maintaining and judging rhythms, Tc = temporal tracking of sounds; TSR = Long-Term Storage Retrieval, Fi = ideational fluency, Fa = associational fluency, Fe = expressional fluency, SM = semantic memory over minutes; aSD = Auditory Sensory Detectors, SPD = auditory immediate memory, Ac = auditory acuity, Va = auditory valence recall; G_s = Clerical Speed; CDS = Correct Decision Speed. From "Remodeling Old Models of Intelligence: Gf-Gc Theory," by J. L. Horn, 1985, in B. B. Wolman (Ed.), *Handbook of Intelligence* (p. 294), New York: Wiley. Copyright 1985 by John Wiley & Sons, Inc. This material is used by permission of John Wiley & Sons, Inc.

Processing speed (Gs)—speed and efficiency in performing automatic or very simple cognitive tasks.

Quantitative knowledge (Gq)—the ability to comprehend quantitative concepts and relationships and to manipulate numerical symbols.

Reading–writing (Grw)—the ability associated with reading and writing; probably includes basic reading and writing skills and the skills required for comprehension and expression.

Academic knowledge (Gak)—the attained level of information and procedures related to scholastic achievements.

Although the broad *Gf* and *Gc* common factors are still prominent aspects of this cognitive system, it is clear that other aspects of cognitive functioning have broad impacts too, including visualization (*Gv*) and audition (*Ga*) and independent aspects of memory (*Glr*, or tertiary storage and retrieval [*TSR*] and *Gsm*, or short-term acquisition and retrieval [*SAR*]) and speediness (*Gs*). These interdependent functions are thought to be important in an understanding of individual differences in cognition over age. The extension of the *Gf-Gc* kinematic-trend predictions is presented for some of these broad factors in Figure 3 (from Horn, 1986). Again, we expect that *Gc* rises over age (with *TSR*) while *Gf* declines with age (with *Gs*, *SAR*, and even *Gv*).

Although there seems to be a broad consensus about major aspects of these age changes in cognitive variables, there is a lack of precision in the descriptions of which cognitive functions change over age. It is still unclear whether the patterns of growth and change differ from one cognitive function to another. The general shape or shapes of the age curves of cognitive functions are still unclear. It is unclear at what chronological age or ages these curves reach their peak and at what age or ages they start to decline. It is also unclear whether the same age curves do, in fact, capture the variation both between persons and within persons. We address these substantive questions in this research.

Recent Longitudinal Methodology

There are many ways to address these substantive questions using empirical data and growth curve models (see McArdle, in press, for a historical overview). Although cross-sectional age studies of persons at one point in time can deal effectively with *age differences* between groups, these studies are incapable of examining *age changes* within a group. For this and other reasons, considerable attention has been directed toward collections of data from persons at multiple points in time (e.g., Collins & Horn, 1991; Nesselroade & Baltes, 1979).

Prior longitudinal studies (e.g., Baltes & Mayer, 1999; Bayley, 1966; Botwinick, 1977; Schaie, 1996) have indicated several methodological problems to keep in mind when interpreting results, including biases that are due to regression toward (and egression away from) the mean, artificial relationships between initial level and change, biases that are due to selective survival, and training, or practice, effects. These studies have also indicated psychological factors that may play a role in the results, including individual differences in attitudes (e.g., about testing), the ease of attending testing sessions, incentives (e.g., pay for volunteering), the conditions under which testing was done (e.g., when in military service, when hospitalized), factors determining attitudes (e.g., the historical period in which participants were raised), ethnicity, geographical region, and many others. Furthermore, the need to make appropriate inferences from models with incomplete longitudinal data is a particularly acute problem for studies of psychological constructs measured over the entire life span.

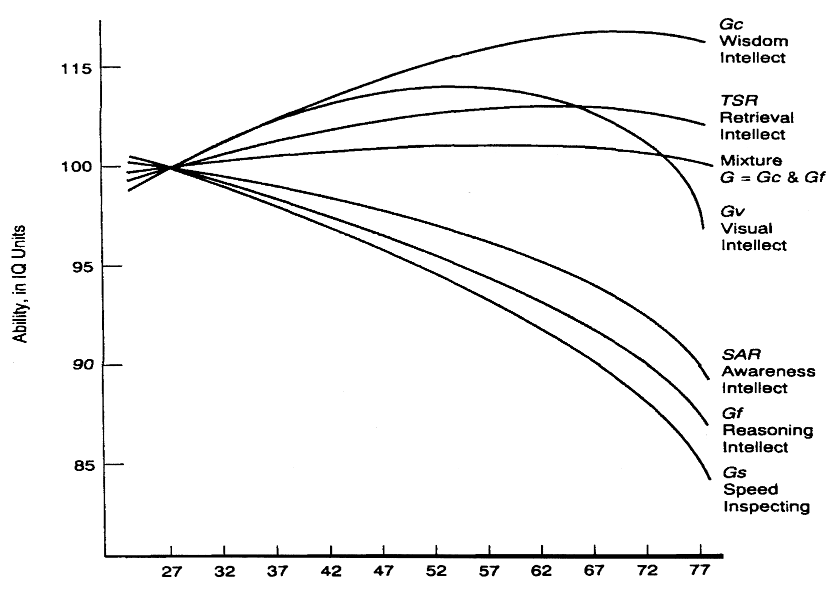

Figure 3. An elaboration of the *Gf-Gc* kinematic-trend hypotheses for the adult life span. From "Intellectual Ability Concepts," by J. L. Horn, 1986, in R. L. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (p. 52), Hillsdale, NJ: Erlbaum. Copyright 1986 by Lawrence Erlbaum Associates. Reprinted with permission.

Some of these methodological problems have been considered in recent statistical techniques in latent change models (see McArdle & Bell, 2000; McArdle & Hamagami, 2001; McArdle & Nesselroade, 1994). These new latent-variable structural equation models (SEMs) follow a traditional growth curve approach. In this article, we write (a) a change model for the true scores or latent variables and (b) an expected trajectory over all times of interest, and (c) we fit the expected trajectories to the available raw score information for each person. Assuming enough data are collected, this approach allows us to (d) fit a model to a time series of data on one or more variables; (e) write a structural model for the means, deviations, and correlations; (f) assume separate individual-difference components for curve characteristics such as levels, slopes, and asymptotes; and, (g) represent growth curve concepts as testable hypotheses.

These newer versions of latent growth models are flexible and include provisions for missing and unbalanced data (McArdle & Anderson, 1990; McArdle & Bell, 2000; McArdle & Hamagami, 1992; McArdle & Woodcock, 1997). Although it is not yet widely recognized, these latent growth SEMs turn out to be mathematically and statistically equivalent to what are popularly known as *random coefficient*, *multilevel*, or *hierarchical* linear models (Bryk & Raudenbush, 1987, 1992; McArdle & Hamagami, 1991, 1992, 1996; Metha & West, 2000; Miyazaki & Raudenbush, 2000; Wishart, 1938). These growth models recognize the multiple-level structure of the data following classical analysis of variance (ANOVA) models—as repeated observations "nested" within individuals, with both "between" and "within" estimation of intercepts and slopes (Metha & West, 2000; Singer, 1998). We take advantage of this parallel development by using some newly available multilevel computer programs to fit some newly available latent growth models (for details, see McArdle & Woodcock, 2000).

Similar multilevel models for time lag have recently been used for cognitive aging research. Sliwinski and Buschke (1999) used longitudinal data to study the impacts of age and processing speed as predictors of other cognitive changes. Wilson, Gilley, Bennett, Beckett, and Evans (2000) used longitudinal time lags to examine the relation of age and other variables to the onset of Alzheimer's disease. Because of the different initial ages of the participants, both studies used longitudinal multilevel analyses of time lags and considered age as a predictor variable in the equations. However, these time-based multilevel analyses are not formally equivalent to using age-based multilevel analyses. That is, these multilevel analyses may provide an accurate representation of time changes (as in McArdle & Woodcock, 1997) but may not be an adequate representation of age changes (as in McArdle & Bell, 2000; McArdle & Hamagami, 1992; McArdle & Woodcock, 2000; Nesselroade & Baltes, 1979; Wohlwill, 1973).

In the present research we use a mixture of age-based and time-based models that employ two time points of data. This accelerated longitudinal design (Aber & McArdle, 1991; Bell, 1954; McArdle & Anderson, 1990; McArdle & Bell, 2000; McArdle & Woodcock, 1997) links all the different age cohorts as they move throughout the two measurement occasions. Two-occasion data permit an initial way to measure individual growth, and data with a wide age spread permit a way to study development over a long period of the life span (e.g., S. C. Duncan & Duncan, 1995; McArdle & Hamagami, 1992; cf. Swanson, 1999). In particular, this approach allows one to fit latent growth models of age using multilevel models that estimate parameters of individual trajectories including intercepts, age slopes, and individual variation in such parameters. These models can also lead to other alternatives, such as planned differences in time lags and incomplete and unbalanced data (see McArdle & Woodcock, 1997, 2000).

Overview of the Current Research

In this study we ask some basic questions such as How do broad cognitive functions grow and change within an individual over age? and Are these growth and change patterns different from one variable to another? We examine two-occasion longitudinal data from a relatively large sample of persons ($N \sim 1{,}200$) given WJ–R tests (McGrew et al., 1991; Woodcock & Johnson, 1989). The WJ–R scales are a wide-range comprehensive set of individually administered tests of intellectual ability, scholastic aptitude, and achievement. At least four features of the WJ–R make it especially valuable as an instrument for research in human development and psychometric change. The WJ–R scales (a) are well-normed from ages 2 to over 90 years, (b) are calibrated using a Rasch model,[1] (c) include multiple ability measures, and (d) can be administered quickly and easily. The equal interval feature of these Rasch-based scales is useful in time-lag research. In theory, differences in the *W* scale can be interpreted to have the same meaning at any performance level, a critical feature of studies in which the change in a score (i.e., a rate) is of primary concern (e.g., see Embretson, 1996; McDonald, 1999; Woodcock, 1999). Despite such psychometric features, the WJ–R has not yet been the focus of a great deal of longitudinal research (McArdle & Woodcock, 1997; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992).

The longitudinal structural equation models we use are all based on the general theme of a latent growth model of age (see Browne & Du Toit, 1991; McArdle, 1988; McArdle & Aber, 1990; McArdle & Epstein, 1987; Meredith & Tisak, 1990). We show how these longitudinal data can be effectively analyzed using contemporary statistical techniques based on latent growth and multilevel models. We emphasize the interpretation of results in terms of developmental components of change, and we elaborate on technical issues in footnotes and appendixes. Our initial models are based on standard polynomial age changes, but we move

[1] *W* scores and Rasch scaling: All WJ–R scales use a constant of 500 arbitrarily set as the mean score of 10-year-olds, and this constant has been removed in most analyses here. Score differences are based on a Rasch model with a logit or log-odds transformation. An important characteristic of any Rasch scale is that a given difference along the scale has the same implication for change in performance at any level and in any area measured. For example, if on retest a person has grown 10 *W* units from the initial score, this person can now perform tasks with 75% success that were performed with 50% success on the initial test (Woodcock, 1999). If the person has declined by 10 units from the initial score, this means the person now performs tasks with 25% success that were formerly performed with 50% success. This relationship is true for any 10-point difference on the *W* scale, whether the test performance is a kindergartner's ability to blend speech sounds or a college student's ability to solve calculus problems.

quickly to nonstandard linear age segments (i.e., splines) models and several nonlinear forms. One of these models, a dual exponential model, has a reasonable substantive basis in aging (e.g., McArdle & Hamagami, 1996; McArdle & Woodcock, 2000). These growth curves are compared across different variables and across different demographic groups. In general, we show how these kinds of longitudinal models are useful for analyses of life span curves from multilevel longitudinal data, and we present some broad substantive implications of these results.

Method

Participants

Between June 1988 and September 1996, we collected longitudinal retest information on persons who had already been measured as part of the norming sample of the WJ–R tests (see McGrew et al., 1991). The present study is based on data from a sample of 1,193 participants who provided WJ–R data on at least two occasions of measurement (for up to 17 WJ–R tests). This sample includes persons ranging in age from 2 to 95 years (median age = 20.3 years, mean age = 26.9 years) who participated in both relatively short-term test–retest (e.g., time lag less than 1 year) and in longer term test–retest (Mdn = 2.2 years, M = 2.7 years). Characteristics of these longitudinal participants are described in Table 1. To simplify this description we separated the sample into five age groups—ages 2–5, 6–10, 11–19, 20–49, and 50–95 years—and into five time-lag groups—time lags of 0–1, 1–2, 2–3, 3–4, and 4–10 years. The sample is not evenly distributed over these groups. The largest subset of people (n = 100) is in the oldest age group (ages 50–95) and was measured over 3–4 years of time. The smallest subset (n = 5) is in the youngest group (ages 2–5) and was measured over the longest time lag (4–10 years). The relationship between age at first test and time lag is relatively small (r = .10).

Table 2 lists the demographic characteristics of the WJ–R norming sample and the retest sample studied here. We selected participants using a randomly stratified sampling designed to create some spread in age and time lag between measurements (based on studies by McArdle, 1994; McArdle & Woodcock, 1997). We started with all persons who had completed the initial testing as part of the norming study (N > 6,000), and we randomly stratified these persons on age, sex, and geographical region over most of the United States. To create the retest sample, we started with a list of the entire norming sample and (a) randomly sorted persons within each stratum into a within-cell list, (b) recontacted persons by mail and phone according to their position on this sorted list, and (c) continued with these contacts until we reached a numerical quota of persons (25% per cell) who agreed to testing. Selection for the second testing was not based on the scores at the first testing. Noncontacts (< 23%) and refusals (< 15%) were lower than the expected ranges for longitudinal studies (e.g., see Schaie, 1996).

Table 1
An Overall Description of the Persons, Ages, and Time Lags Sampled in the WJ–R Study

Time lag between test and retest	Age (in years) at first testing					
	2–5	6–10	11–19	20–49	50–95	All ages
Less than 1 year						
n	43	24	77	65	35	244
% of total	3.6	2.0	6.5	5.5	2.9	20.5
Median age (years) at Test 1	4.0	8.0	18.0	23.0	74.0	19.0
Median time lag (years)	0.6	0.8	0.8	0.7	0.6	0.7
1–2 years						
n	43	82	101	68	40	334
% of total	3.6	6.7	8.5	5.7	3.4	28
Median age (years) at Test 1	4.0	8.0	14.0	31.0	67.5	14.0
Median time lag (years)	1.3	1.3	1.3	1.3	1.25	1.3
2–3 years						
n	38	11	10	25	23	107
% of total	3.2	0.9	0.8	2.1	1.9	9
Median age (years) at Test 1	2.0	6.0	16.5	31.0	67.0	16.0
Median time lag (years)	2.6	2.6	2.7	2.7	2.8	2.7
3–4 years						
n	34	46	57	95	100	332
% of total	2.8	3.9	4.8	8.0	8.4	27.8
Median age (years) at Test 1	4.0	7.5	14.0	36.0	65.0	31.0
Median time lag (years)	3.4	3.6	3.4	3.4	3.3	3.4
4–10 years						
n	5	25	51	66	29	176
% of total	0.4	2.1	4.3	5.5	2.4	14.8
Median age (years) at Test 1	5.0	8.0	16.0	33.0	66.0	21.0
Median time lag (years)	4.2	4.2	8.0	7.0	7.0	7.0
All time lags						
n	163	188	296	319	227	1,193
% of total	13.7	15.8	24.8	26.7	19	100
Median age (years) at Test 1	4.0	8.0	16.0	31.0	67.0	20.3
Median time lag (years)	1.8	1.6	1.4	3.0	3.1	2.2

Note. All persons listed here participated in at least two testing sessions using some of the same WJ–R tests. WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised.

Table 2
Demographic Characteristics of the Retest Sample and the Norming Sample

Characteristic	Statistic	Retest sample (N = 1,193)	Norming sample (N = 6,471)
Age (years)	M (SD)	27.4 (22.9)	20.3 (18.2)
	Minimum	2.00	1.92
	Maximum	95.1	95.6
Gender			
Male	n (%)	555 (46.5)	3,152 (48.7)
Female	n (%)	638 (53.5)	3,317 (51.3)
Ethnicity			
White non-Hispanic	n (%)	797 (68.8)	4,445 (68.7)
Black non-Hispanic	n (%)	207 (17.4)	1,054 (16.3)
American Indian	n (%)	19 (1.6)	71 (1.1)
Asian Pacific	n (%)	60 (5.0)	197 (3.0)
Hispanic	n (%)	110 (9.2)	592 (9.1)
Missing	n (%)	—	112 (1.7)
Education (years)	M (SD)	10.4 (4.8)	9.11 (4.9)
	Minimum	0	0
	Maximum	21.0	25.0
Educational attainment			
No high school	n (%)	491 (41.2)	3,440 (53.2)
High school	n (%)	163 (13.7)	519 (8.0)
No college	n (%)	264 (22.1)	1,183 (18.3)
College	n (%)	50 (4.2)	189 (2.9)
Beyond college	n (%)	81 (6.8)	322 (5.0)
Missing	n (%)	144 (12.1)	818 (12.6)
Occupational level			
White collar	n (%)	214 (17.9)	714 (11.0)
Blue collar	n (%)	147 (12.3)	463 (7.2)
Service	n (%)	83 (7.0)	230 (3.6)
N/A	n (%)	749 (62.8)	4,952 (76.6)
Occupational status			
Employed	n (%)	244 (20.5)	776 (12.0)
Unemployed	n (%)	34 (2.8)	156 (2.4)
Not in labor force	n (%)	190 (15.9)	561 (8.7)
N/A	n (%)	725 (60.8)	4,866 (75.2)

Note. N/A (not applicable) includes school-age participants, housewives, and dropouts.

The retest participants were selected to reflect a range of age, sex, and geographical regions. As can be seen in Table 2, these participants were reasonably similar to the norming sample on several other characteristics. Although the longitudinal sample was slightly older (by design), most other demographic characteristics were the same as those for the larger WJ–R norming samples. An interview was arranged for each person at a convenient and quiet location (school, home, public library, etc.), and after informed consent was established, each person was examined by one interviewer for approximately 3 hr.

Variables Measured

Table 3 is a list and description of the WJ–R scales that were used in this study. Included are the 14 basic measures of the WJ–R that are intended to measure seven broad factors used in contemporary theories of intellectual ability theory: Fluid Reasoning (*Gf*), Comprehension–Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Short-Term Memory (*Gsm*), Processing Speed (*Gs*), Auditory Processing (*Ga*), and Visual Processing (*Gv*). These first seven scales are each created as an unweighted average of two indicators (listed in Table 3), and each is reported to have relatively high internal consistency reliability ($r_{ic} \sim .90$; see Appendix C of McGrew et al., 1991). Three additional clusters are widely used in academic and other settings and are also presented in Table 3: Broad Quantitative Ability (*Gq*), Broad Academic Knowledge (*Gk*), and Broad Reading and Writing (*Grw*). These three clusters are based on multiple measures with high internal consistency reliability. The last variable, Broad Cognitive Ability (*BCA*; see McGrew et al., 1991), is derived by averaging the first seven composite scores.

It may also be useful to note that many of the participants were also measured on a variety of other intellectual abilities (e.g., the Wechsler Adult Intelligence Scale, Cattell's Matrices, Power Letter Series) as well as given a comprehensive demographic questionnaire, but these variables are not discussed here. However, in order to keep the overall testing time at the second occasion to about 3 hr, only 17 of the 22 WJ–R measures were administered to all participants. That is, for the most part, this incomplete data structure was planned in advance of the data collection to permit the measurement of a wide range of variables (after McArdle, 1994). However, because of the small longitudinal sample sizes for some variables (e.g., Picture Vocabulary with n = 327), and in order to keep the data structure as simple as possible, several factors are represented here on the basis of only one variable (e.g., *Gc* is simply Oral Vocabulary). Because of the relatively high reliability, all composites are treated in the same way in further analyses here.

Data Description

Table 4 lists the means, standard deviations, and minimum and maximum scores for each of the 11 WJ–R composites for the retest sample used in this study and for the overall WJ–R norming sample (N = 6,471) from

Table 3
A Description of the WJ–R Variables Used in the Test–Retest Study

WJ–R factor cluster	Internal consistency (r_{ic})	WJ–R test	$N[1]/N[2]$	Internal consistency (r_{ic})
Fluid Reasoning (*Gf*)	.946	Analysis–Synthesis (AS)	1,044/1,090	AS = .900
		Concept Formation (CF)	1,045/1,087	CF = .941
Comprehension-Knowledge (*Gc*)	.936	Picture Vocabulary (PV)	1,193/327	PV = .879
		Oral Vocabulary (OV)	1,050/1,098	OV = .900
Long-Term Retrieval (*Glr*)	.945	Memory for Names (MN)	1,193/1,192	MN = .916
		Visual Auditory Learning (VAL)	1,193/285	VAL = .909
Short-Term Memory (*Gsm*)	.890	Memory for Words (MW)	1,071/1,070	MW = .800
		Memory for Sentences (MS)	1,193/1,191	MS = .865
Processing Speed (*Gs*)	.866	Visual Matching (VM)	1,047/1,091	VM = .799
		Cross Out (COU)	1,047/1,040	COU = .752
Auditory Processing (*Ga*)	.888	Incomplete Words (IW)	1,193/1,191	IW = .787
		Sound Blending (SB)	1,193/268	SB = .867
Visual Processing (*Gv*)	.816	Visual Closure (VC)	1,193/1,192	VC = .721
		Picture Recognition (PR)	1,193/254	PR = .808
Broad Quantitative Ability (*Gq*)	.954	Calculation (CA)	1,193/263	CA = .925
		Applied Problems (AP)	1,192/1,106	AP = .920
Broad Academic Knowledge (*Gak*)	.949	Science (SC)	1,193/1,146	SC = .875
		Social Studies (SS)	1,193/1,141	SS = .887
		Humanities (HU)	1,193/1,133	HU = .877
Broad Reading and Writing (*Grw*)	.943	Letter Word Identification (LWI)	1,193/1,081	LWI = .936
		Passage Completion (PSC)	1,049/1,050	PSC = .895
		Dictation (DI)	1,193/1,140	DI = .908
Broad Cognitive Ability (BCA)	.948	Average of first seven factor composites	1,193/1,044	

Note. Sample sizes ($N[1]$ = test sample; $N[2]$ = retest sample) are based on available data. WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised. All reliabilities are from Appendix C in *Woodcock–Johnson Revised Technical Manual* (pp. 265–282), by K. S. McGrew, J. K. Werder, and R. W. Woodcock, 1991, Itasca, IL: Riverside Publishing. Copyright 1991 by Riverside Publishing Company. Reprinted with permission.

which this longitudinal sample was drawn. The differences found here between the two samples are all very small, with the retest sample scoring, in general, about one to two points (*W* units) lower. A multivariate statistical test on two independent samples (logistic prediction of the test–retest sample vs. the norm minus test–retest sample) shows that these group differences account for a trivial amount ($R^2 < 0.01\%$) of the overall variation in the data.

Table 5 is a more complete description of the raw data collected in this test–retest study for the 11 WJ–R composites for the five age groups. It is important to again note that both test and retest scores obtained on each of the scales were originally collected on a Rasch-based measurement scale (see Footnote 1). This further implies that differences or rates of change between scale points can be meaningfully compared both within and between variables.

To convey the key features of the distributions of these scores, we present the various change statistics within each cell of Table 5. The first two rows of each cell provide a simple indication of the rate of change in scores from one testing time to the next (for calculation, see Table 6, Equation 1). The first row is the median slope or change score (50th percentile), and the second row includes values from both the lower quartile (25th percentile) and an upper quartile (75th percentile). These values can be interpreted as an expected score difference for each year of time (e.g., a linear slope). For example, the median slope score obtained on the BCA composite for all participants (lowest cells) is 50% = 2.1, with a range of 25% = 0.4 and 75% = 5.2. The third row in each cell is the standard test–retest correlation, uncorrected for either age or time lag. This simple relationship between the BCA scores, $r_{[1,2]}$, is .89 over all ages and times.

In general, most WJ–R scores obtained here show that the younger age groups (ages 6–10 and 11–19) exhibit lower starting scores but more positive score changes, whereas the older age groups (ages 20–49 and 50–95) display higher initial scores and less change. In most columns of Table 5, there are systematic changes in scores related to age differences. Some variables exhibit much larger age changes than others (e.g., Comprehension–Knowledge, Broad Academic Knowledge, Broad Reading and Writing). Also, although we have not provided details here, these 1-year slope estimates seem to get smaller as the time lag between tests increases (e.g., across the last row). The test–retest correlations exhibit a substantial amount of variation but exhibit a slight tendency to be higher with shorter time lags. The first age column also shows that only a few of the youngest children (ages 2–5) were measured on all seven factors (leading to the reduced *n*), but in further analyses we included any available data. These 1-year change estimates and their ranges may be most useful when used as descriptive indicators of the expected range of changes for persons of specific ages on specific tests.

Table 4
Descriptive Statistics for the WJ–R Cognitive Variables in the Longitudinal Sample and in the WJ–R Norming Sample

WJ–R factor cluster	Retest sample (N = 1,193)	Norming sample (N = 6,471)
Fluid Reasoning		
M (*SD*)	506.2 (19.9)	508.6 (19.4)
Minimum	435.5	433.0
Maximum	553.5	553.5
Comprehension–Knowledge		
M (*SD*)	536.6 (20.2)	537.0 (20.3)
Minimum	448.0	433.5
Maximum	575.5	595.0
Long-Term Retrieval		
M (*SD*)	497.3 (13.0)	499.7 (13.8)
Minimum	447.0	434.0
Maximum	537.5	537.5
Short-Term Memory		
M (*SD*)	510.1 (18.6)	512.6 (19.1)
Minimum	441.5	398.0
Maximum	566.5	573.5
Processing Speed		
M (*SD*)	513.9 (18.2)	516.4 (18.3)
Minimum	443.5	417.0
Maximum	563.0	563.0
Auditory Processing		
M (*SD*)	501.3 (17.6)	503.3 (17.0)
Minimum	433.0	422.0
Maximum	540.0	552.5
Visual Processing		
M (*SD*)	504.9 (12.7)	506.7 (12.7)
Minimum	454.0	430.0
Maximum	533.5	538.5
Broad Quantitative Ability		
M (*SD*)	532.8 (22.1)	534.6 (23.2)
Minimum	437.5	386.0
Maximum	601.5	607.5
Broad Academic Knowledge		
M (*SD*)	529.3 (18.2)	530.4 (18.7)
Minimum	448.0	368.3
Maximum	571.7	574.3
Broad Reading and Writing		
M (SD)	529.1 (18.0)	529.0 (17.9)
Minimum	421.0	421.0
Maximum	570.0	570.5
Broad Cognitive Ability		
M (SD)	510.1 (13.9)	512.2 (13.6)
Minimum	456.0	440.0
Maximum	540.0	540.0

Note. WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised.

Figure 4 includes several plots of the Broad Cognitive Ability (BCA) variable for all persons having complete data at both occasions (n = 1,044). We have subtracted the constant 500 (the empirical average score at age 10) from each score for ease of presentation and modeling. Figure 4A is a scatter plot of the BCA scores for each person at the first (x-axis) and second (y-axis) occasions of testing; this plot illustrates the generally high test–retest correlation. Figure 4B is a scatter plot of the initial age at testing (x-axis) versus the BCA scores at the first occasion of testing (y-axis); this plot shows a distinct curvature in the trend *between individuals* at different ages—there is a steep rise up to about age 20 and then a gradual decline afterward. Figure 4C is a scatter plot of the age at testing (x-axis) versus the 1-year changes (see Equation 1 in Table 6) in BCA scores (y-axis); here we see larger positive changes *within individuals* at the earlier ages and a different pattern of changes at later ages. Finally, Figure 4D is an individual line plot of the BCA scores that includes two ages of testing and two BCA scores; here we see the same general curvature but now with the combined cross-sectional (Figure 4D) and longitudinal (Figure 4C) information.

Linear Growth and Change Models

We now consider a few kinds of structural models for the changes over time that can be applied to the available WJ–R data, and these are outlined in Table 6. In the notation used here, the longitudinal data collection is based on the measurement of a group of persons (n = 1 to N) at two occasions (m = 1 or 2); the observed score at each occasion, $W_{m,n}$; and the corresponding chronological ages, $Age_{m,n}$.

The first equation (Equation 1) in Table 6 defines an observed rate-of-change score (symbolized as W) as the W-score difference ($W_2 - W_1$) divided by the amount of time between the test and the retest ($Age_2 - Age_1$). There are many published data analyses in which these kinds of observed rates-of-change scores are outcomes in further data analyses (e.g., Giambra, Arenberg, Zonderman, Kawas, & Costa, 1995; McCrae, Arenberg, & Costa, 1987; Sullivan, Rosenbloom, Lim, & Pfefferman, 2000). Unfortunately, the simple and practical use of rates-of-change calculations has also led to a great deal of theoretical complexity and confusion. Prior methodological research has shown that when observed rates are used as outcomes in multiple regression analyses, the resulting coefficients can be severely biased by several factors, including residual error, measurement error, and regression (and egression) to the mean (e.g., Allison, 1990; Nesselroade & Bartsch, 1977; Nesselroade & Cable, 1974; Nesselroade, Stigler, & Baltes, 1980; Raykov, 1999; Willett, 1990; Williams & Zimmerman, 1996). These problems can be severe when standard linear regression is used with time-dependent variables (e.g., Boker & McArdle, 1995; Hamagami & McArdle, 2001; McArdle, Hamagami, Elias, & Robbins, 1991).

Methodological solutions to these problems have been proposed in recent statistical techniques in latent growth and change models (see McArdle & Bell, 2000; McArdle & Hamagami, 2001; McArdle & Nesselroade, 1994; Meredith & Tisak, 1990). In the approach used here, we estimate the parameters of a model for the latent change score by specifying an expected set of latent trajectories over time. In Equation 2 of Table 6 we define (a) an initial starting score ($w[0]$) at a specific starting time (t = 0); (b) a latent slope score (Δw), defined as the theoretical difference in a pair of latent scores ($w[t]$ and $w[t - \Delta t]$); and (c) an error of measurement score at each occasion of measurement (e_m).

A path diagram including additional aspects of the linear latent growth model is presented in Figure 5. In the model used here we assume the unobserved initial-level component ($w[0]$) has mean and variance (i.e., μ_0 and σ_0^2), and the error of measurement has a mean of zero, has constant variance ($\sigma_e^2 > 0$), and is uncorrelated with every other component. Here, however, the constant change component (Δw) has a nonzero mean (i.e., μ_Δ, the average of the latent change scores), a nonzero variance (i.e., σ_Δ^2, the variability of the latent change scores), and a nonzero correlation with the latent initial levels (i.e., $\rho_{0\Delta}$). From these standard assumptions, we can write a more complex set of expectations for the means, variances, and covariances for all observed scores. This approach to the estimation of latent parameters allows us to avoid some of the key problems raised by observed change scores (e.g., see Cnaan, Laird, & Slasor, 1997; Goldstein, 1995; Lindsey, 1993; Littell, Miliken, Stoup, & Wolfinger, 1996; Pinherio & Bates, 2000; Singer, 1998; Verbeke & Molenberghs, 2000).

The test–retest data collection used here includes some variation with respect to the time lag between tests (e.g., see Table 1). The models described earlier have not made full use of this variation, and a variety of alternative models are possible. For example, if we consider that some aspects of the observed score may be due to prior familiarity with the test

Table 5
One-Year Rates of Changes by Age Group for 11 WJ–R Factor Composites

WJ–R factor composite	Change statistic	Age (in years) 2–5	6–10	11–19	20–49	50–95	All ages
Fluid Reasoning	Median Δ	8.7	5.1	3.0	2.0	1.2	3.3
(*Gf*; *n* = 1,192)	Range of Δ	5.0–13.8	2.4–8.4	0.2–7.1	0.0–5.6	−0.2–4.3	0.4–7.7
	Test–retest *r*	.63	.65	.65	.66	.80	.85
Comprehension-Knowledge	Median Δ	8.7	6.0	2.6	1.4	0.3	2.3
(*Gc*; *n* = 1,046)	Range of Δ	6.2–11.3	2.5–8.9	0.0–5.7	0.0–3.8	−1.9–3.0	0.0–6.0
	Test–retest *r*	.49	.74	.86	.84	.91	.90
Long-Term Retrieval	Median Δ	6.0	3.4	2.6	0.6	−0.3	1.6
(*Glr*; *n* = 1,192)	Range of Δ	2.7–11.2	0.0–8.3	0.0–12.0	−1.4–6.2	−2.4–1.6	−0.8–7.7
	Test–retest *r*	.54	.68	.60	.64	.58	.70
Short-Term Memory	Median Δ	9.2	3.9	1.6	0.7	0.3	2.1
(*Gsm*; *n* = 1,191)	Range of Δ	4.1–13.6	1.4–9.0	−1.4–6.0	−1.6–5.3	−2.3–4.2	−1.1–7.6
	Test–retest *r*	.65	.78	.77	.77	.66	.85
Processing Speed	Median Δ	7.8	6.6	2.4	1.2	0.3	2.1
(*Gs*; *n* = 1,042)	Range of Δ	7.1–12.7	3.9–9.0	0.4–5.2	−0.4–3.5	−1.6–2.5	−0.1–5.6
	Test–retest *r*	.30	.71	.75	.84	.86	.85
Auditory Processing	Median Δ	6.8	5.0	0.6	0.5	0.5	0.6
(*Ga*; *n* = 1,191)	Range of Δ	0.0–12.0	0.0–10.3	−2.2–7.7	−2.7–6.8	−2.3–6.8	−2.1–8.1
	Test–retest *r*	.60	.78	.57	.58	.65	.82
Visual Processing	Median Δ	8.7	5.1	3.0	2.0	1.2	3.3
(*Gv*; *n* = 1,192)	Range of Δ	5.0–13.8	2.4–8.4	0.2–7.1	0.0–5.6	−0.2–4.3	0.4–7.7
	Test–retest *r*	.63	.65	.65	.66	.80	.85
Broad Quantitative Ability	Median Δ	15.7	9.1	0.3	0.0	0.0	3.0
(*Gq*; *n* = 1,106)	Range of Δ	7.6–23.5	4.3–13.3	−3.0–7.0	−3.0–4.5	−2.9–4.2	−1.4–10.3
	Test–retest *r*	.58	.73	.83	.82	.82	.92
Broad Academic Knowledge	Median Δ	10.9	6.5	2.6	1.0	−0.1	2.7
(*Gak*; *n* = 1,141)	Range of Δ	7.1–15.2	4.3–8.4	−0.3–5.2	−0.5–3.5	−2.3–1.6	−0.2–7.6
	Test–retest *r*	.70	.86	.90	.88	.91	.94
Broad Reading and Writing	Median Δ	22.1	8.4	2.4	0.4	−0.2	1.7
(*Grw*; *n* = 961)	Range of Δ	16.3–26.2	5.3–13.9	−0.1–5.8	−1.3–3.1	−1.8–1.6	−0.7–7.1
	Test–retest *r*	.29	.78	.92	.89	.94	.90
Broad Cognitive Ability	Median Δ	5.6	4.9	2.6	1.2	0.3	2.1
(BCA; *n* = 1,033)	Range of Δ	5.2–7.4	3.2–7.0	1.0–5.3	0.2–4.0	−0.8–2.2	0.4–5.2
	Test–retest *r*	.61	.83	.88	.87	.90	.89
Sample sizes used		163–161	188–177	269–295	280–319	218–227	1,193–995

Note. WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised.

itself, we can define a set of models representing retest, or practice, components (e.g., McArdle & Woodcock, 1997), as presented in Equation 3 of Table 6.

Here the standard linear model applies to the first time of testing, but the observed score at the second time of testing has a constant addition (p). Although this practice component is unobserved, we assume it is a feature of the test and not the trait (i.e., it only appears after the first occasion of testing) and that it is the same at any interval of time. In the simplification of this model illustrated in Figure 5, we assume this practice has a mean (i.e., the average practice μ_p) and a variance (i.e., the variability of practice σ_p^2) and no correlation with the latent initial levels or the latent changes. A variety of more complex models are possible to use here.[2]

Nonlinear Growth and Change Models

The simple linear models described above are not always a reasonable representation for longitudinal growth data (see, e.g., Figures 1 and 3). This situation has led researchers to use more complex models, often based on a polynomial series (Bryk & Raudenbush, 1987, 1992; Cohen & Cohen,

[2] Alternative linear models: In the linear-change-plus-practice model (Equation 3 of Table 6), we allow linear changes (Δp) in the size of the practice effect with increasing time lags ($Age_2 - Age_1$) between the test and retest (see McArdle & Woodcock, 1997). With the use of this same

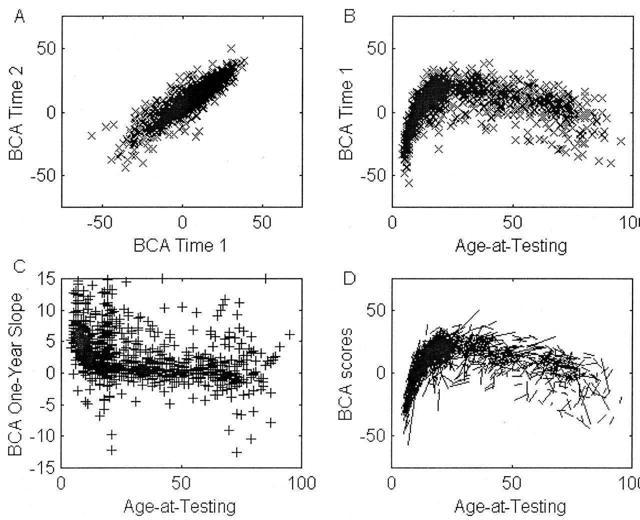

Figure 4. A description of the Woodcock–Johnson Psycho-Educational Battery—Revised (Woodcock & Johnson, 1989) Broad Cognitive Ability (BCA) longitudinal data (retest $n = 1{,}044$).

1983; Goldstein, 1995; Stimson, Carmines, & Zeller, 1978) and written as in Equation 4 of Table 6, in which four components of change are introduced (e.g., $\Delta w1$, $\Delta w2$, $\Delta w3$, and $\Delta w4$). In this type of model, the latent difference score is assumed to change as a function of age, so the resulting trajectory over age forms polynomial shapes with multiple points of inflection. Of course, there are several well-known problems with fitting polynomials to longitudinal data, including (a) a lack of direct interpretation of the coefficients associated with each polynomial power and (b) the need for further often arbitrary constraints to deal with two-occasion data (McArdle & Woodcock, 1997).[3]

Our key goal in this analysis is to find a structural model that (a) captures the apparent nonlinearity of the growth curve and (b) requires only a small number of meaningful individual-difference components. One of the simplest alternative models we can consider here is based on the joining of "connected segments" of linear growth (see Bryk & Raudenbush, 1992; Cudeck, 1996; Draper & Smith, 1981; McArdle, Paskus, & Boker, in press; Smith, 1979). This kind of SEM can be written as Equation 5 in Table 6 and is based on the definition of the intercept ($w[0]$) at some specific "knot point" or cutoff age (C), a constant slope score (Δb) for the person "before" this age ($Age < C$), and a second constant slope (Δa) "after" at all later ages ($Age > C$). The observed score at any specific age is thus assumed to be an additive accumulation of the changes that have occurred up to that age, and at the specific location of the cutoff points, the segments are "joined." These age-segmented SEMs can also be fitted using the multilevel software described in Appendix A. In contrast to the polynomial restrictions used above, this approach seems to be a reasonable way to ensure the identification of the other model parameters. This use of linear age segments can be expanded to include multiple cutoffs (C_i) between any specific ages. For example, a model including five segments can be defined by the four age-group cutoffs of Tables 1 and 5 (i.e., ages 6, 11, 20, and 50; Draper & Smith, 1981).

The previous multicomponent models are useful when particular age or time segments are well defined. There also exist a variety of more complex functional forms in which the components of the model can be used to estimate these kinds of critical ages. Meredith and Tisak (1990) showed how previous research could be reinterpreted as a structural equation model written in the form of a latent curve with estimated basis coefficients (i.e., $B[t] = \beta[t]$). In recent work, McArdle and Hamagami (1992, 1996) and McArdle and Bell (2000) have shown how this kind of a latent growth model with an explicit structure on these latent loadings can be useful for models of age-sampled test–retest data.

In the analyses presented here, we consider a latent growth model based on a concept of competing forces (after Cerella & Hale, 1994; McArdle, 2001; Sandland & McGilchrist, 1979; Simonton, 1984; Zeger & Harlow, 1987), which is written as Equation 6 of Table 6. In this model, we assume there are two rates (π_b and π_a) representing an accumulative rise before ($\pi_b > 0$), and an accumulative fall after ($\pi_a > 0$), some unobserved but estimable cutoff age ($t = \tau_b$). These two exponential impacts are assumed to be related to two different sources of individual differences (b and a), but for the purposes of identification in test–retest data, these are aggregated into one latent slope score ($\Delta w[\tau_b, \tau_a]$). The difference between these two exponential accumulations at any specific age is used as the loading ($\beta[t]$) for the age-related trajectory. The model parameters define other aspects of these growth functions, including the expected age (τ_a) of the latent maximum score (i.e., the peak) and the expected age (τ_b) of the latent asymptote score (i.e., the initial decline; for details, see McArdle & Woodcock, 2000).

Comparing Growth and Decline Curves

One overall theme in this research is the necessity of some procedure for comparing the growth patterns between different variables. Typically, they are compared by using a multivariate analysis of variance with a test for what are often termed *parallel growth curves* across different variables

logical distinction it is possible to include more complex models for the practice scores ($p[t]_n$), including a set of models with interactions of practice effects with age. However, these kinds of models now include four (or more) individual-differences terms. Because of the basic limits of information available in test–retest data, we often cannot estimate all variance terms simultaneously (examples to follow in the text). Given these kinds of constraints, other researchers have found it useful to place constraints so the models represent the change from a baseline with initial age held constant (Sliwinski & Buschke, 1999; Verbeke & Molenberghs, 2000; Wilson et al., 2000). These restricted but popular mixed models may be useful for understanding the time-lag components—but they are clearly not focused on age components, so they are not examined further here (for further details, see McArdle & Woodcock, 2000).

[3] Identification restrictions for two-occasion data: The use of two-occasion longitudinal data places some limits on the number of unique parameters that can be estimated. This situation leads to numerical results that show an overparameterization of the model, odd estimates (correlations approaching 1 or −1), no unique computation of the standard errors, and the need for further and often arbitrary constraints (see McArdle & Woodcock, 1997). For example, in the higher order polynomial models (see Equation 4 of Table 6), the mean coefficients are all unique under the assumption that a minimal set of variance components ($p < 3$) is estimated. In the latent growth spline model, each age segment is allowed to have its own change means (μ_b and μ_a) and change variances (σ_b^2 and σ_a^2) and possibly some correlation among components (σ_{ba}, for persons overlapping the cutoff age). In order to allow an optimal fit for the scores, the levels and slopes in such a model are allowed to correlate (i.e., σ_{0b}, σ_{0a}). However, because most of the data collected will be in one age segment or the other, it is appropriate to carry out estimation with a restriction of no correlation among the age slopes across segments (i.e., $\sigma_{ba} = 0$, or based on the overlapping persons). As we fit an increasing number of age segments with only two-occasion data, the covariance constraints become necessary for identification (i.e., all $\rho_{ba} = 0$).

Table 6
Algebraic Expressions of the Latent Growth Models Fitted to the WJ–R Data

Model	Equation
Linear growth and change models	
Observed rate-of-change model	(1) $\Delta W_n/\Delta t = (W_{2,n} - W_{1,n})/(Age_{2,n} - \text{Age}_{1,n})$, where $W_{m,n}$ = observed score on person n at measurement m, and $Age_{m,n}$ = observed age of person n at measurement m.
Latent change score model	(2) $W_{1,n} = \{w[0]_n + Age_1 \cdot \Delta w_n\} + e_{1,n}$, and $W_{2,n} = \{w[0]_n + Age_2 \cdot \Delta w_n\} + e_{2,n}$, where $w[0]_n$ = latent initial-level score of person n, Δw_n = latent slope score of person n, and $e_{m,n}$ = latent error score of person n at measurement m.
Retest, or practice, components model	(3) $W_{1,n} = \{w[0]_n + Age_1 \cdot \Delta w_n\} + e_{1,n}$, and $W_{2,n} = \{w[0]_n + Age_2 \cdot \Delta w_n\} + p[0]_n + (Age_2 - Age_1) \cdot \Delta p_n + e_{2,n}$, where $p[0]_n$ = latent initial practice score of person n, and Δp_n = latent slope of practice score of person n.
Nonlinear growth and change models	
Polynomial growth model	(4) $W_{m,n} = w0_n + Age_m \cdot \Delta w1_n + Age_m^2 \cdot \Delta w2_n + Age_m^3 \cdot \Delta w3_n + Age_m^4 \cdot \Delta w4_n + e_{m,n}$, where Age_{mn}^p = age basis of power p for person n, and Δwp_n = latent polynomial component score of person n.
Connected linear spline model	(5) $W_{m,n} = w[0]_n + B[Age_m < C] \cdot \Delta b_n + B[Age_m > C] \cdot \Delta a_n + e_{m,n}$, where C = the latent age cutoff score (estimated or fixed), Δb_n = latent slope "before" the age cutoff for person n, and Δa_n = latent slope "after" the age cutoff for person n.
Growth model of "competing" forces	(6) $W_{m,n} = w[0]_n + \beta[Age_m] \cdot \Delta w[\tau_b,\tau_a]_n + e_{m,n}$ with $\beta[t] = \exp(-\pi_b \cdot Age_m) - \exp(-\pi_a \cdot Age_m)$, where $\beta[t]$ = the accumulation of a latent age basis, π_b = latent rate "before" the age peak, π_a = latent rate "after" the age peak, and $\Delta w[\tau_b,\tau_a]_n$ = the combined latent slope for person n.
Multivariate models to compare growth and decline curves	
SEM enhanced "parallel growth curves" model	(7) $Y[Age_m]_n = \mu_y + \lambda_y \cdot g[Age_m]_n + u_y[Age_m]_n$, and $X[Age_m]_n = \mu_x + \lambda_x \cdot g[Age_m]_n + u_x[Age_m]_n$, with $g[Age_m]_n = g[0]_n + \beta[Age_m] \cdot \Delta g_n + e[Age_m]_n$, where λ_q = the latent loading (proportion) for variable q, and μ_q = the latent intercept (additive) for variable q.

Note. WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised. SEM = structural equation modeling.

(Bock, 1975; Pothoff & Roy, 1964; Rao, 1958; Rencher, 1995; Zuker, Zerbe, & Wu, 1995). In these analyses, we assume multiple observed scores Y and X are independently measured over ages and times, and we write a model based on Equation 7 of Table 6. In this model, both measured variables (Y and X) are related within each time to a third latent variable (g) by a common factor model with variable intercepts (μ_y, μ_x), factor loadings (λ_y, λ_x), and uniquenesses (u_y, u_x). Over time we assume that only the common factor score ($g[Age_m]$) changes as a function of one of the basis coefficients ($\beta[Age_m]$), as described above (i.e., linear or nonlinear). This SEM is consistent with the idea that each measured variable is only an indicator of a latent construct with factorial invariance over age and time (as in McArdle, 1988; McArdle & Nesselroade, 1994; McArdle & Woodcock, 1997). A path diagram of this general structural growth model is presented here in Figure 6.

In this bivariate representation of the common factor model, we need to add additional constraints to ensure that the parameters are identified. This results in a model in which the two growth curves are assumed to be proportional (λ) to one another except for constants representing the differences in origin (μ_x) and scale ($u_x[Age_m]$). By comparing the fit of this combined bivariate model to that of the separate models, we obtain a formal test of the fit of the "same shape" hypothesis. Obviously, this is only one of many SEMs for describing the similarities of the shape of the growth curves across multiple variables, but it turns out to be both simple to fit and useful.

Maximum Likelihood Estimation of Longitudinal Multilevel Models

One benefit of the formal model expressions in Table 6 is that they can be fitted directly to data and examined as empirical alternatives. Each of these alternative models yields an explicit set of expected means, variances (standard deviations), and correlations over time. By comparing these expected statistics to the observed statistics from the data, we can (a) estimate an optimal set of values for the unknown parameters and (b) examine the goodness of fit of the model. To the degree that the alternative

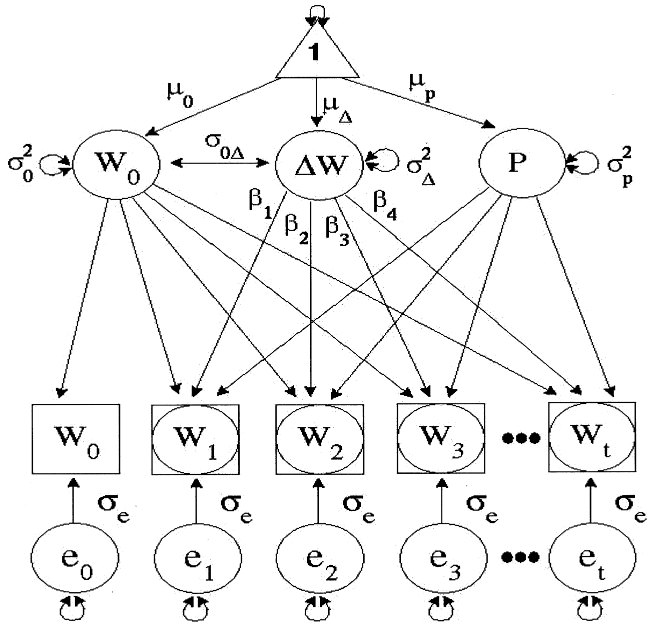

Figure 5. A path diagram of the multilevel growth model. W_t = score at time t; ΔW = slope; P = practice; e = uniqueness; β_n = basis coefficient; σ_0^2 = variance of level; σ_Δ^2 = variance of slope; $\sigma_{0\Delta}$ = covariance of level and slope; σ_p^2 = variance of practice; μ_0 = mean of level; μ_Δ = mean of slope; μ_p = mean of practice.

models yield different expectations, we have the possibility of detecting empirically grounded differences between these theoretical concepts.

The computer software needed for these kinds of multilevel growth model analyses is widely available (for reviews, see Bryk & Raudenbush, 1992; Hox & Kreft, 1994; Jöreskog & Sörbom, 1999; Kreft, De Leeuw, & van der Leeden, 1994). Each of these computer programs provides maximum likelihood estimates (MLEs) and standard-likelihood-based statistical information on goodness of fit (e.g., $f = -2\text{LL}$ and chi-square) for a wide variety of what are now commonly termed *multilevel* or *random coefficient* models (Bryk & Raudenbush, 1987, 1992; McArdle & Hamagami, 1996). Although there are some practical differences among these computer programs, results from the PROC MIXED and PROC NLMIXED algorithms of SAS are presented here (e.g., Cnaan et al., 1997; Littell et al., 1996; Singer, 1998; for detailed scripts, see McArdle & Woodcock, 2000, and Appendix A in the present article). We fit these structural (i.e., multilevel) models using all available data with a likelihood function (*f*) that is typically based on multivariate normal theory (i.e., $f = -2\text{LL}$; Little & Rubin, 1987; McArdle, 1988, 1994), and a variety of goodness-of-fit indices are used.[4,5]

Results

Four sets of numerical results are presented here in summary form: (a) a summary of growth and decline models based on polynomial multilevel strategies; (b) details on models with linear age splines (see Table 7); (c) elaboration of nonlinear models based on dual competition (see Tables 8 and 9); and (d) a summary of differences between growth curves for different abilities using common factor models.

Initial Multilevel Polynomial Models

The multilevel analyses were initiated using a standard sequence of linear polynomials of increasing complexity, and we can illustrate these results by describing the numerical results for the BCA variable (see Figure 4). A first model fit included only a single mean (μ_0 = 7.9) and a single error variance ($\sigma^2_{e\{0\}}$ = 233.4) and yielded a baseline likelihood (f = 17,626). A second model included individual differences in initial levels (σ_0^2 = 206.6), reduced the size of the error variance (σ_e^2 = 36.3), and improved the fit, $\chi^2(1)$ = 1,245, η^2 = .844. The next model fit was a linear growth model (see Equation 2 in Table 6) with an intercept mean (μ_0 = 11.8) at age 20 and a slope mean ($\mu_\Delta = \mu_1$ = 0.294) for

[4] Goodness of fit: All of the models we examine represent different hypotheses about the growth and decline of cognitive abilities. In our analyses, we compare these models and select the most accurate results on both statistical (i.e., statistically significant) and conceptual (i.e., prior research) grounds. As an overall index of goodness of fit we calculated the difference in the likelihood ($f\{j - k\}$) for nested models. Under normal theory, this statistic is assumed to be distributed as a noncentral chi-square variate with degrees of freedom based on the differences in the number of nested parameters ($df\{j - k\}$). As a second index of fit we calculated a likelihood increment percentage (LIP; LRT = likelihood ratio test; i.e., $\text{LIP}\{j\} = 1 - [\text{LRT}_{\text{model}\{j\}}/\text{LRT}_{\text{baseline}}] \times 100$), also popular in other contexts (see Horn & McArdle, 1980; Menard, 2000). We also calculated the percentage reduction in error (PRE), documented by Snijders and Bosker (1994). These percentages are assumed to be unbiased estimates of the modeled variance at the first level (PRE_1) and at the second level (PRE_2) for a correctly specified population, so we used these as one indicator of the impact of specific predictor variables (for details, see Kreft & De Leeuw, 1998; Singer, 1998; Snijders & Bosker, 1994). These analyses include multiple statistical tests, and because we did not use any adjustment in our analyses, we cannot claim to enjoy any overall experimentwise protection. In a standard multiple-model comparison, we could presume we had an overall null hypothesis and then adjust the alpha level by, say, using a traditional Bonferroni correction procedure. However, as soon as we are interested in another analysis, which we usually are, we would need to go back and correct the prior analyses for these additionally new tests. As far as we know, this problem of the likely overstatement of results by multiple and continual testing is not solved, and about all we can offer is to report our statistical indices in the context of mathematical modeling for others to use and openly critique. Instead, we identify results for those parameters that were not twice as large as their own standard errors (i.e., $p\ [H_0] > .05$; see Tables 7 and 8).

[5] Fitting multivariate multilevel models: Recent research has shown how any bivariate model (see Table 6, Equation 7) can be reparameterized and fit using the available mixed-model software (Goldstein, 1995; McArdle, Paskus, & Boker, in press). The observed data for $Y[Age]$ and $X[Age]$ can be lined up as a single vector ($W[Age]' = [Y[Age]' \mid X[Age]']$), and by making appropriate provisions for the expectations of the model for each variable, the standard single-outcome optimization can treat this multiple variable model as a single equation. It follows that it is possible to construct a simultaneous test of proportionality of multiple ($M > 2$) growth variables, estimating the proportions (λ_j, λ_k, etc.) as factor loadings and obtaining an overall fit to test the hypothesis of a general growth factor (as in McArdle, 1988; McArdle & Woodcock, 1997). In practice, however, these kinds of multiple-outcome variable models turn out to be complex models to fit (McArdle et al., in press).

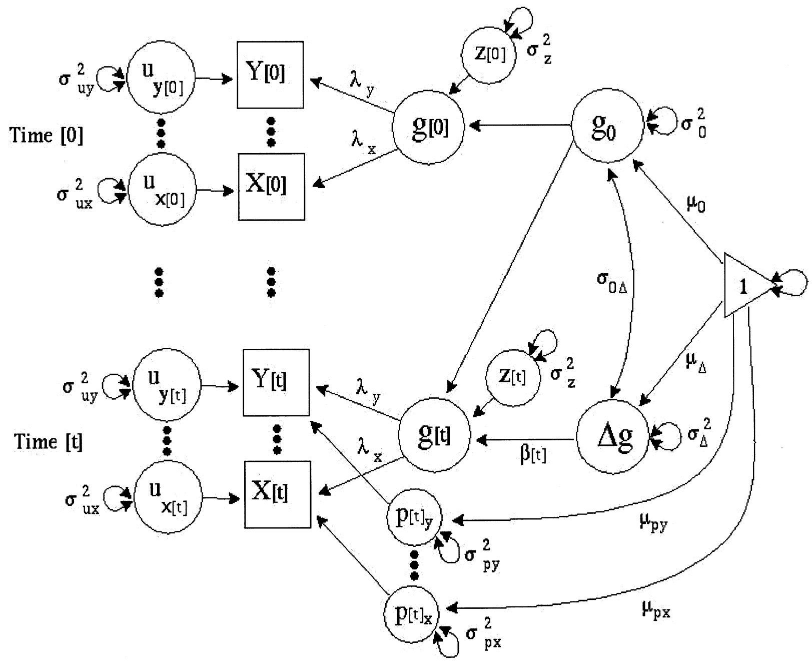

Figure 6. A path diagram of the multivariate multilevel growth model. X[t] = score of variable X at time t; Y[t] = score of variable Y at time t; u = uniqueness; g[t] = general factor at time t; Δg = slope of general factor; g_0 = initial level of general factor; z[t] = disturbance of factor g at time t; p = practice component; β = basis coefficient; μ = means of components; σ^2 = variances of components; λ = loading of variable on factor g.

each year of age before and after age 20. The variance of the age-related changes (σ_1^2 = .464, ρ_{01} = −.64) was notable; there was a decrease in the size of the first-level residual variance (σ_e^2 = 28.4) and a corresponding improvement in the model fit, $\chi^2(3)$ = 253.

The next three models fit to BCA added successive polynomial growth models to the means. Because we only have two time points of data for each person, we limited the estimation of variance components to the intercept and slope but estimated all means of the higher order powers. The fit of each progressive higher order polynomial component yielded a significant increase in the fit of the model: second-order quadratic model, $\chi^2(1)$ = 253, $\Delta\eta^2$ = .034; third-order cubic model, $\chi^2(2)$ = 1,529, $\Delta\eta^2$ = .072; and fourth-order quartic model (Equation 3 of Table 6), $\chi^2(3)$ = 1,805, $\Delta\eta^2$ = .083. Higher order analyses of these models of the means were not assessed further here. This last model seems to represent an important increase in fit, and the expected mean trajectory over age from this fourth-order polynomial model is plotted in Figure 7A. The particular parameters (see Appendix B) are not clearly separable as components of the resulting curve, but the general picture (Figure 7A) clearly shows a rapid increase in the early years, followed by a peak and a slower decline over all adult years.

The numerical sequence results for all other WJ–R variables are essentially similar to those presented above. The estimated parameters are different for each variable, but the general shapes of the curves are at least as complex as the one presented above. For example, the estimated fourth-order curves for *Gf* and *Gc* variables obtained reasonable fits (f = 17,126 and 17,209, respectively), and when plotted, these polynomial curves have the same general shape as the curve in Figure 7A; however, the parameters are different, which leads to important differences in the curves (as in Figure 1). These growth curve differences are best described using the other models that follow.

Linear Age-Segmented Multilevel Results

The next models fitted were based on a linear growth model with two nested age segments—before and after the cutoff age of

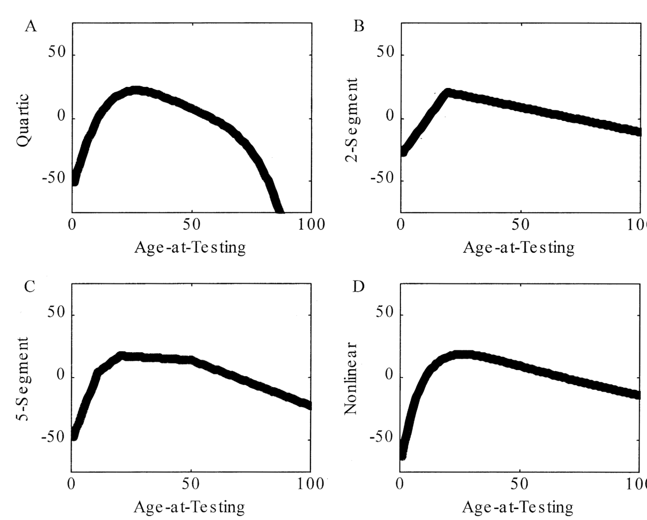

Figure 7. Alternative linear and nonlinear multilevel models fitted to Broad Cognitive Ability longitudinal data.

$C = 20$ years. Table 7 is a list of results from the parameters of this kind of a bilinear latent growth model fitted to each of the 11 WJ–R composites over all ages. Each row represents one WJ–R variable, and each column represents a particular kind of parameter: the mean and variance of the intercept (column 1), the mean and variance of the "before age 20" segment (column 2), the mean and variance of the "after age 20" segment (column 3), the mean and variance of the independent practice shift (column 4), the variance of the independent error component (column 5), the difference in fit ($\Delta\chi^2$) between this model and the baseline model (column 6), and the difference in modeled variance ($\Delta\eta^2$; column 7).

This bilinear model for the BCA scores yields an intercept mean ($\mu_0 = 20.6$) at age 20, a positive pre-20 slope mean ($\mu_b = 2.64$) for each year before age 20, and a smaller and negative slope mean ($\mu_a = -0.39$) for each year after age 20. Now the expected 10-year age change from age 10 to age 20 is +26.4, whereas the expected 10-year change from age 20 to age 30 is −3.9 points (in the *W* scales). The key difference here is the separation of the large yearly increases before age 20 and the small yearly decreases after age 20. In this bilinear model, the age-related variance changes are also split into two independent components ($\sigma_b^2 = 1.35$, $\sigma_a^2 = .05$) with the before-age-20 changes showing the larger yearly component. In addition, this model adds a practice effect (coded as a 0 at the first test and a 1 at the retest). The parameters obtained show an average score increase that is due to practice ($\mu_p = 3.2$) but no estimable individual differences in this shift ($\sigma_p^2 = 0$ with correlations fitted as zero for identification). The addition of two age segments plus a practice effect does improve the model fit a great deal compared with either the constant baseline model, $\chi^2(8) = 1{,}857$, the linear growth model, $\chi^2(5) = 1{,}660$, or the bilinear without-practice model, $\chi^2(2) = 319$. Furthermore, there is a decrease in the size of the first-level residuals ($\sigma_c^2 = 13.6$), yielding an additional explained variance ($\Delta\eta^2_{\{7-1\}} = .10$). The practice mean is added at one time only and accounts for slightly more mean differences than do the overall trait changes seen over any 1 year of age. Additional forms of practice effects (e.g., linear decline with time, different practice mean before and after age 20)

Table 7
Linear Latent Growth Model Parameters Including Bilinear Age and Practice Components for WJ–R Variables

WJ–R composite	Age 20 constant μ_0 (σ_0^2)	Ages 2–19 segment μ_1 (σ_1^2)	Ages 21–75 segment μ_2 (σ_2^2)	Additive practice μ_p(σ_p^2)	Random error (σ_e^2)	Fit of changes $\Delta\chi^2$	Modeled variance η^2 ($\Delta\eta^2$)
Fluid Reasoning	16.9 (98.1)	2.68 (1.90)	−0.45 (0.086)	6.21 (1.23)	(47.0)	1,988	.882 (.118)
Comprehension–Knowledge	41.4 (329.5)	5.11 (3.45)	−0.08 (0.308)	1.81 (4.4)	(45.5)	2,631	.947 (.066)
Long-Term Retrieval	9.1 (131.3)	1.52 (1.07)	−0.45 (0.02?)	3.6 (3.59)	(57.2)	927	.743 (.086)
Short-Term Memory	22.4 (259.0)	3.38 (3.90)	−0.36 (0.071)	2.1 (≈0?)	(83.3)	1,368	.868 (.061)
Processing Speed	31.6 (151.7)	3.95 (2.70)	−0.62 (0.126)	2.7 (≈0?)	(36.1)	1,272	.923 (.099)
Auditory Processing	11.2 (58.3)	2.07 (3.01)	−0.34 (0.045)	1.8 (≈0?)	(64.6)	1,427	.820 (.041)
Visual Processing	16.9 (98.1)	2.68 (1.89)	−0.45 (0.086)	6.2 (1.2?)	(47.0)	1,989	.882 (.118)
Broad Quantitative Ability	49.8 (632.3)	6.64 (12.60)	−0.54 (0.551)	2.8 (≈0?)	(114.3)	2,466	.945 (.066)
Broad Academic Knowledge	43.2 (384.4)	5.38 (5.08)	−0.46 (0.281)	1.9 (9.2)	(33.7)	2,593	.970 (.063)
Broad Reading and Writing	43.0 (587.3)	5.51 (17.10)	−0.44 (0.350)	1.9 (60.8)	(12.4)	2,125	.989 (.124)
Broad Cognitive Ability	20.6 (86.1)	2.64 (1.35)	−0.39 (0.05)	3.2 (~0?)	(36.3)	1,857	.942 (.098)

Note. All parameters are maximum likelihood estimates fitted using SAS PROC MIXED and LISREL 8.3. A question mark (?) next to a value indicates a parameter that is not twice as large as its own standard error (i.e., $p[H_0] > .05$). WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised.

were also fit, but the results are not listed here because they did not add anything to the fit or alter the parameter estimates. The bilinear equation for the BCA scores is plotted over two age segments (purged of a practice mean of 3.2) in Figure 7B.

The previous model estimates suggest that each of the different WJ–R scales has some notable, systematic age differences in the mean and variation of age changes. The pattern of mean changes is very similar for all 11 variables: (a) large yearly mean increases and substantial change variance over the age range from 2 to 19 years, (b) smaller yearly mean decreases and smaller change variance from ages 21 to 95, and (c) positive mean shifts for this simple form of practice. A relatively large proportion of the variance is due to the initial level of the trait (ranging from about .70 to .90), and these levels differed across tests. Individual differences in the linear age segments accounted for a small but potentially important part of the trait variance (from .03 to .10). But there are some important differences apparent here as well. For example, the estimated equations for *Gf* and *Gc* show that the intercepts and variances of these equations are quite different, and the key age-related feature is clear—the *Gc* rises more rapidly up to age 20 (5.11 vs. 2.68), and then *Gf* falls more rapidly (−0.45 vs. −0.08).

More complex spline models were fit using a five-nested-age-segments model with the cutoffs $C_1 = 6$, $C_2 = 11$, $C_3 = 20$, and $C_4 = 50$. The cutoffs were chosen to produce segments matching the five age groups defined earlier (in Tables 1 and 5). These models permit five different mean and variance components to be estimated (only the first two are listed) and also include the practice component for any time lag. In the first case, we obtained means that clearly showed more systematic growth in the early years followed by declines in the later years (per year average change, $\mu_\Delta = 5.24, 4.92, 1.46, -0.12$, and -0.73); there was also a positive practice mean ($\mu_p = 2.75$). In this case, the estimation of the restricted variance components was not completely successful, but the resulting models did fit better than the previous two-segmented models, $\chi^2(8) = 259$, but with only a small increase in overall fit ($\Delta\eta^2_{\{8-7\}} = .004$). These nonlinear trends were not matched with estimated variances as in a single model, but it is likely that this model was computationally limited (e.g., by multiple correlation parameters from two-occasion data and by the 1-year scaling of age). The expected trajectory over age from this five-segment model is plotted in Figure 7C.

Nonlinear Multilevel Results for Broad Cognitive Ability

Table 8 expands these issues by fitting nonlinear models to the same WJ–R data. Here we centered all data (at age = 0) so time would constantly increase. The key model reported here is a dual exponential model (of Equation 6) with both level and slope variance and covariance.

The final nonlinear estimates for BCA are plotted in Figure 7D. The parameters of this model have direct substantive interpretation (see Appendix B). The mean of the initial level ($\mu_0 = -75.2$, with scores centered at a constant of 500) is the initial score (at age = 0), and the mean slope ($\mu_\Delta = 117.7$) is the constant multiplier on the changing age basis. The initial growth rate ($\pi_b = .0065$, so $\exp\{-.0065\} = .994$) is strong and positive, and the decline rate ($\pi_a = .1165$, so $\exp\{-.1165\} = .890$) shows a steady but slower decline. From these parameters of the model we can deduce that the maximum BCA average score, is $\mu_{w[\tau b]} = 18.3$ (see Table 9), and this occurs at an age of $\tau_b = 26.2$ years (i.e., where the first derivative equals zero). We can also deduce that the point where the slope begins to de-accelerate is at a score of $\mu_{w[\tau a]} = 8.0$ and an age of $\tau_a = 52.3$ years (i.e., where the second derivative equals zero; Raudenbush & Chan, 1993). The expected rates of changes at specific ages can also be calculated, and here we find the expected average rate of change in BCA between ages 2 and 19 is $\mu_s = 4.0$ *W* units per year, whereas between ages 20 and 75 it is only −0.3 *W* units per year.

The latent variance components can be interpreted in many ways as well. The initial-level variance (at age 0, $\sigma_0^2 = 91.7$), the slope ($\sigma_\Delta^2 = .628$), and the covariance ($\sigma_{0\Delta} = -7.5$) can be combined to form expectations about the latent or true score variance at any age, and the error variance ($\sigma_e^2 = 13.1$) can be added to form the total variance. For BCA, at the peak age of $\tau_b = 26.2$ years, the true score variance is $\sigma^2_{w[26]} = 73.9$. These formulas can be combined with the expressions for the mean to produce 95% confidence boundaries around the latent scores. A pictorial representation of these variance expectations is presented in Figure 8. It is noteworthy that these variances do not exhibit much "fan-spread" (Cook & Campbell, 1979). This figure also includes the individual line plots of the longitudinal data to illustrate the goodness of fit of this dual exponential with-practice model.

For BCA, this model yielded a much better fit to the data than found before, $\chi^2(5) = 1{,}838$, $\Delta\eta^2 = .09$, so we explored this model in further detail. In more restrictive models (not shown in Table 8) we found that we could eliminate the slope variance, $\chi^2(2) = 1$, but we could not eliminate the level variance, $\chi^2(2) = 50$. In an expanded form of this model we added a practice component (coded as 0 or 1 depending on the occasion of testing), which yielded a substantial improvement in fit, $\chi^2(7) = 2{,}070$, $\Delta\eta^2 = .08$. In the final model, we added an exponential loss function (to mimic a proportional practice decay), but the addition of this parameter ($\pi_p = -.028$) did not yield an improvement in fit. Statistical comparisons among these last few models suggested that (a) the overall double exponential model is a much better fit than the other models used here; (b) either level or slope components are needed, but the level is more important; and (c) the inclusion of a practice mean is helpful to model fit.

The final nonlinear estimates for *Gf* and *Gc* have several different features, including (a) growth rate differences, with *Gf* slowing more rapidly than *Gc* (.0052 vs. .0026), and (b) decline rate differences, with *Gf* falling more rapidly than *Gc* (.1539 vs. .1104). The resulting parameters of these nonlinear models can be drawn as smooth latent curves (with 95% confidence boundaries) for any variable. In Figure 9 we give a complete picture of the *Gf* and *Gc* nonlinear model expected trajectories and observed data.

In Figures 10 (*Gf*, *Gc*, *Gs* and *Glr*) and 11 (*Gsm*, *Ga*, *Gv* and *Gq*), we include expectations for all eight composites without the specific data. In Table 9 we list additional statistical information for all 11 WJ–R composites, including the expected (a) age at peak growth (τ_b), (b) age at initial decline (τ_a), (c) latent means and

Table 8
Nonlinear Latent Growth Model Parameters Including Bilinear Age and Practice Components for WJ–R Variables

	Time-lag test and age model mean and variance components							
	Age = 0 constant	Age slope	Growth rate	Decline rate	Practice + covariance			Modeled variance
WJ–R composite	μ_0 (σ_0^2)	μ_1 (σ_1^2)	π_b [exp]	π_a [exp]	μ_p (σ_{01})	Random error (σ_e^2)	Fit of changes $\Delta\chi^2$	η^2 ($\Delta\eta^2$)
Fluid Reasoning	−116.5	156.3	.0052	.1539	3.0		3,003	.831
	(110.9)	(16.8)	[.995]	[.857]	(30.5)	(67.5)		(.067)
Comprehension-Knowledge	−116.3	179.2	.0026	.1104	≈0?		2,631	.943
	(292.4)	(4.72)	[.997]	[.896]	(4.7)	(49.0)		(.062)
Long-Term Retrieval	−56.7	75.0	.0080	.1805	2.7		1,122	.743
	(16.6)	(82.6)	[.992]	[.835]	(7.3)	(57.2)		(.086)
Short-Term Memory	−101.1	132.7	.0038	.1578	1.8		1,675	.879
	(196.0)	(0.281)	[.996]	[.854]	(−7.2)	(76.4)		(.072)
Processing Speed	−128.3	188.9	.0058	.1295	2.3		2,132	.925
	(140.1)	(0.668)	[.994]	[.879]	(−9.7)	(34.9)		(.101)
Auditory Processing	−84.3	107.0	.0045	.1628	≈0?		1,581	.817
	(88.5)	(2.26)	[.996]	[.850]	(−14.1)	(65.8)		(.038)
Visual Processing	−74.6	113.4	.0075	.1210	5.0		2,283	.884
	(93.5)	(0.973)	[.993]	[.886]	(−9.5)	(46.4)		(.120)
Broad Quantitative Ability	−192.7	254.0	.0022	.1475	≈0?		3,071	.964
	(145.5)	(40.5)	[.998]	[.863]	(75.8)	(100.0)		(.085)
Broad Academic Knowledge	−120.1	179.8	.0035	.1225	0.3		3,082	.970
	(120.4)	(30.9)	[.997]	[.885]	(52.2)	(34.1)		(.063)
Broad Reading and Writing	−398.6	438.0	.0005	.2407	−0.4		2,699	.965
	(189.2)	(16.0)	[.999]	[.786]	(53.9)	(39.2)		(.100)
Broad Cognitive Ability	−75.2	117.7	.0065	.1165	2.7		2,070	.944
	(91.7)	(0.628)	[.994]	[.890]	(−7.5)	(13.1)		(.100)

Note. All parameters are maximum likelihood estimates fitted using SAS PROC MIXED and LISREL 8.3. A question mark (?) next to a value indicates a parameter that is not twice as large as its own standard error (i.e., p [H_0] > .05). WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised.

deviations at these ages, and (d) the implied growth and decline rates of change at the predefined age ranges of 2–19 years and 20–75 years.

Statistical Differences Between Growth Patterns

There are several obvious differences between the WJ–R composites that are noteworthy. For example, when we compare *Gf* with *Gc* (see Figure 9), we see that *Gf* has a slightly slower initial growth rate ($\exp\{-\pi_b\}$ = .995 vs. .997) but a faster decline ($\exp\{-\pi_a\}$ = .857 vs. .896). These parameters reflect an earlier peak age for *Gf* than for *Gc* (τ_b = 22.8 vs. 35.6 years) and an earlier age at decline (τ_a = 45.5 vs. 71.3 years). These results are quite consistent with the basic age-curve predictions of the theory of fluid and crystallized intelligence (Cattell, 1971; Horn & Cattell, 1967; see Figures 1 and 3).

These statistical differences in growth and decline curves further suggest that different rates of change are to be expected for different ages and for different functions, and these results can be used in planning further experiments. For example, Table 9 lists an expected rate of change for *Gf* as 5.5 for ages 2–19 but −0.5 for ages 20–75. This means that the same magnitude of change is expected when we measure the growth of *Gf* in a child over 1 year and the decline of *Gf* in an adult over 11 years. Table 9 also lists an expected rate of change for *Gc* as 7.0 for ages 2–19 but as −0.01 for ages 20–75. This means that the same changes are expected when we measure the growth of *Gc* in a child over 1 year and the decline of *Gc* in an adult over all years. These general age functions suggest we will find little or no change at all if we try to measure cognitive abilities right around their peak ages. These results further suggest that different experimental time lags and different ages will be optimal for measuring changes in people on different constructs.

Most of the other composites studied here are more similar in shape to this *Gf* composite (Analysis–Synthesis + Concept Formation) than to this *Gc* score (only Oral Vocabulary). Among the first seven composites, the fastest growth and decline rates are

Table 9
Implied Characteristics of Nonlinear Latent Growth Curves for All WJ–R Cognitive Composites

	Time-lag test and age model mean and variance components					
	Age at peak	Age at deceleration	Score at peak	Score at deceleration	Rate of Δ ages 2–19	Rate of Δ ages 20–75
WJ–R composite	τ_b (β_{tb})	τ_a (β_{ta})	μ_b (σ_b^2)	μ_a (σ_a^2)	μ_1 (σ_1^2)	μ_2 (σ_2^2)
Fluid Reasoning	22.8 (0.80)	45.5 (0.79)	17.6 (13.3)	6.7 (13.0)	5.5 (8.0)	−0.5 (8.2)
Comprehension-Knowledge	35.6 (0.92)	71.3 (0.87)	44.5 (18.5)	35.7 (17.9)	7.0 (17.7)	−0.01 (17.8)
Long-Term Retrieval	18.1 (0.83)	36.1 (0.75)	6.2 (9.2)	0.10 (8.6)	2.4 (5.9)	−0.4 (5.5)
Short-Term Memory	24.2 (0.89)	48.4 (0.83)	17.1 (13.5)	9.3 (13.6)	4.8 (8.7)	−0.3 (8.5)
Processing Speed	25.1 (0.83)	50.2 (0.75)	27.7 (11.2)	12.6 (11.1)	6.6 (7.2)	−0.6 (7.1)
Auditory Processing	22.7 (0.88)	45.4 (0.82)	9.7 (8.1)	2.9 (8.2)	3.8 (5.4)	−0.3 (5.2)
Visual Processing	24.5 (0.78)	49.1 (0.69)	14.0 (8.9)	3.8 (9.0)	3.8 (5.8)	−0.4 (5.7)
Broad Quantitative Ability	29.0 (0.92)	57.9 (0.88)	42.1 (17.9)	30.9 (17.6)	9.7 (10.6)	−0.3 (11.1)
Broad Academic Knowledge	29.8 (0.87)	59.7 (0.81)	37.2 (15.3)	25.6 (15.0)	6.6 (9.1)	−0.3 (9.4)
Broad Reading and Writing	25.8 (0.98)	51.6 (0.97)	33.2 (17.5)	28.4 (17.4)	15.2 (11.1)	−0.2 (11.6)
Broad Cognitive Ability	26.2 (0.80)	52.3 (0.71)	18.3 (8.6)	8.0 (9.0)	4.0 (5.8)	−0.3 (5.7)

Note. All entries are calculated from maximum likelihood estimates and fitted using SAS PROC NLMIXED. The dual exponential model yielded loadings of $\beta_t = [\exp(-\pi_b * Age_t) - \exp(-\pi_b * Age_t)]$. WJ–R = Woodcock–Johnson Psycho-Educational Battery—Revised.

found for Long-Term Retrieval (*Glr*), which exhibits unexpected peaks at the earliest ages (τ_b = 18.1 years, τ_a = 36.1 years), whereas the slowest growth and change is seen for *Gc*. Perhaps not too surprisingly, the three composites reflecting academically related information (e.g., *Gq*, *Gk* and *Grw*; the last two are not plotted) have much faster growth rates and slower decline rates than the first seven cognitive abilities.

There are several techniques we can use to study the relationships *among* the variables. Some recent models described in the statistical literature have emphasized the examination of parallel growth curves, including the correlation of various components (McArdle, 1988, 1990; Willett & Sayer, 1994). Of substantive interest here is the possibility of examining the equality of the shapes of the group curves (e.g., $B_y[t] = B_x[t]$). Other models can be fitted to examine the size and sign of the covariances of initial levels (i.e., $|\sigma_{y0,x0}| > 0$) and the covariance of slopes (i.e., $|\sigma_{ys,xs}| > 0$). These random coefficients reflect individual similarities in the way persons start and change over time across different variables, and these are key features to some researchers (e.g., T. E. Duncan, Duncan, Strycker, Li, & Alpert, 1999; Raykov, 1999; Willet & Sayer, 1994). It has also been demonstrated how more restrictive hypotheses about proportional growth curves are the expectations from a model in which multiple variables are reflections of the same common factor (e.g., McArdle, 1988, 2001; McArdle & Woodcock, 1997). Such a model can be fitted by including common factor scores ($z[t]$), proportionality via factor loadings (λ_y, λ_x), and uniqueness (u_y, u_x). If multiple measurements are made, this common factor hypothesis about the change pattern is a strongly rejectable model. To formalize these differences and the similarities of the overall growth patterns, we fit the one-common-factor models to specific pairs of variables (see Equation 7 in Table 6; see also Footnote 5 and Appendix B).

In a structural analysis we tested the hypothesis posed in Figure 1: How different are the life span growth patterns of *Gf* and *Gc*? To

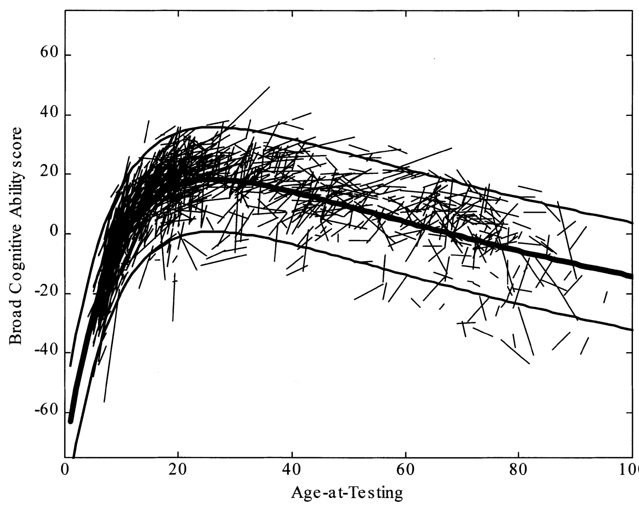

Figure 8. Broad Cognitive Ability longitudinal data and nonlinear multilevel model expectations.

answer this question, we fitted a bivariate model in which the *Gf* scores were presumed to have the same basic curve as the *Gc* scores except for scaling. This new 11-parameter model resulted in a constant of proportionality (λ = 1.55) and new rate parameters (π_b = .0053 and π_a = .1085). However, this bivariate model, when compared with the univariate likelihoods found for each separate variable, yielded a large loss of fit, $\chi^2(7)$ = 1,016. So, while there is distinct similarity in the general growth and decline shapes of *Gf* and *Gc*, this statistical evidence supports the notions that the peaks of growth and decline in *Gf* and *Gc* occur at notably different ages and that the general latent growth patterns are not likely to come from the same source (i.e., *g*).

The same bivariate or multiple-outcome multilevel analysis can be fitted to each of the composites listed here. In order to simplify and limit the potential models, we report only models comparing each of the first seven factors with the BCA overall composite. As a simple summary of results from these analyses we found the following: *Gf* with λ = 1.24 and $\chi^2(7)$ = 169; *Gc* with λ = 1.66 and $\chi^2(7)$ = 472; *Glr* with λ = 0.56 and $\chi^2(7)$ = 747; *Gsm* with λ = 1.14 and $\chi^2(7)$ = 198; *Gs* with λ = 1.39 and $\chi^2(7)$ = 152; *Ga* with λ = 0.80 and $\chi^2(7)$ = 232; *Gv* with λ = 0.90 and $\chi^2(7)$ = 448; and *Gq* with λ = 2.11 and $\chi^2(7)$ = 340. These proportions (λ) can be rescaled to approximate factor loadings for a single-factor model estimated under fixed factor score assumptions (McDonald, 1999).

The goodness-of-fit indices listed above imply that there is a relatively large distance between the overall BCA curve and the seven WJ–R composites. Thus, further approximation to the common factor loadings is not crucial because it appears that these separate growth curves reflect different cognitive functions over age. As found in earlier work on this topic (e.g., McArdle, 1988; McArdle & Woodcock, 1997; and many others), this developmental evidence suggests that a single common factor will not account for all the growth and change among these variables.

Cross-Validation Analysis

There are many ways in which the results presented here can be examined with standard techniques of cross-validation or sensitivity analyses. In a typical kind of analysis we can randomly split the sample in half, fit the same models, and examine the differences in results; that is, what is the overall accuracy if we apply expectations from one model to the data of the other set? In this approach, we gain information about the likely prediction error. Alternatively, we can select the subsamples representing different substantive groupings to see how local the expectations can be; that is, what if we had sampled a particular age group? In this approach, we expect completely different parameters depending on the range of ages. Combinations of these cross-validation techniques can be found in more extensive bootstrap-type analyses that use all the available data.

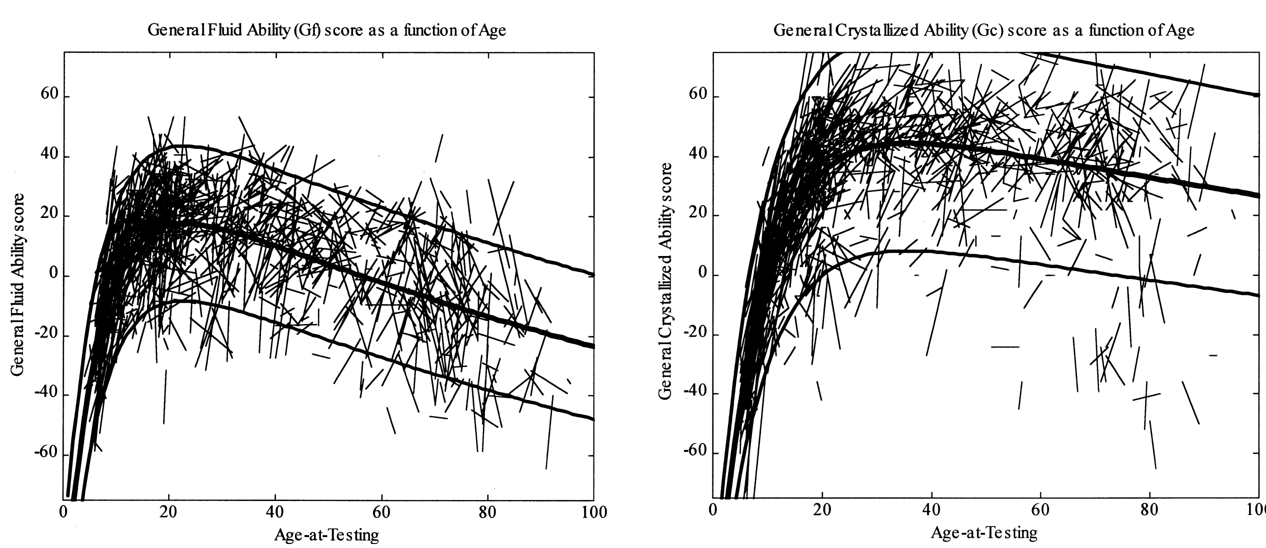

Figure 9. A comparison of multilevel longitudinal age curves for *Gf* and *Gc* abilities (Rasch *W* units).

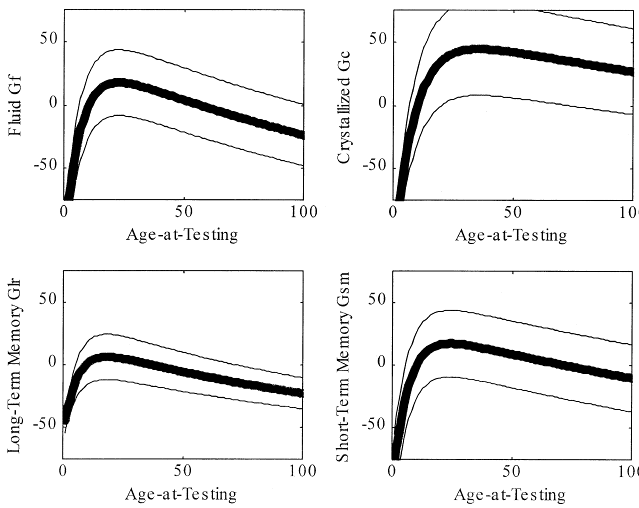

Figure 10. Expected latent curves for *Gf*, *Gc*, *Glr*, and *Gsm*.

We fitted the exponential growth and decline curves to the BCA composite using data from the first and second occasions separately. We then compared the results with the estimates from the overall sample. These analyses yielded some notable differences in parameter values from the first and second and overall assessments—mean intercepts ($\mu_0 = -81.8$ and -79.3), age slopes ($\mu_1 = 117.3$ and 120.7); growth rates ($\pi_b = .0052$ and .0065), decline rates ($\pi_a = .1352$ and .1153), and error variance ($\sigma_e^2 = 93.4$ and 93.8). All of these parameters were accurately estimated (t values > 2.0). This simple form of this cross-validation analysis gives a positive estimate of the accuracy of the values obtained when all the available data are used, and the other variables exhibited similar statistical behavior.

Several parameter estimates in the overall model cannot be uniquely estimated using one data point only: the variance and covariance terms and the mean of practice at the second occasion. To examine the adequacy of these parameters, we examined what happened when progressively larger numbers of data points were not included (i.e., as if they were not measured in the study). We carried this out for random subsamplings in which 20%–80% of the data were eliminated. In these analyses, the parameter estimates for the intercept and slope variances (σ_0^2 and σ_1^2) remained fairly stable up to about an 80% loss, but other parameters (e.g., the intercept–slope covariance, σ_{01}) were not as easy to replicate. However, and in general, the parameters representing the means retained their overall value over different samplings of persons with this data set. A more extensive bootstrap-type analysis using all of the available data may be needed for improved information about the probable range of error in these data.

Discussion

Summary of Results

The main substantive results of this study suggest that it is possible to understand the growth and decline of most broad cognitive functions in terms of a family of curves that rise and fall (after Horn, 1970). Most clearly, as shown in Figures 1 and 3, the functions describable as broad fluid reasoning (*Gf*) and acculturated crystallized knowledge (*Gc*) are separable entities that have different growth patterns. The same result seems to follow for different kinds of broad memory (*Glr* and *Gsm*), processing speed (*Gs*), and auditory and visual processing (*Ga* and *Gv*) and for several forms of academic knowledge (*Gq*, *Grw*, and *Gk*).

We mainly used the BCA score as a convenient starting point to highlight our overall modeling strategy and to compare with other variables (e.g., see Figure 1). But there are reasonable scientific concerns about the singular and often inappropriate use of such a composite score. Many prior analyses have shown that these kinds of scores lack internal validity when development is considered (e.g., Horn & Cattell, 1966; McArdle & Woodcock, 1997). This BCA variable is often considered as an indicator of *g,* or general intelligence (from Spearman, 1904; see Jensen, 1998), but the results presented here, and in concert with prior theory about fluid and crystallized intelligence, suggest that the description of a cognitive system with only a single *g* factor is an overly simplistic view of the more complex sequential dynamics (as in Devlin, Daniels, & Roeder, 1997; E. B. Hunt, 1995; McArdle & Woodcock, 1997, 1998; cf. Humphreys, 1989; Jensen, 1998). This does not mean that the BCA score cannot be useful in other contexts—the BCA is a simple unweighted composite of more fundamental constructs. As such it may yield some benefits in external validity (prediction) in the same way as any compensatory equation (e.g., high school grades and tests; see McArdle, 1998). However, the analyses presented here describe BCA, compare it with other abilities, and indicate that it does not have the validity required of a developmental construct (e.g., McArdle & Prescott, 1992).

This means that the use of a single general construct has limited developmental construct validity. For example, some prior biometric genetic research by Plomin, Pedersen, Lichtenstein, and McClearn (1994) demonstrated a differential heritability of general cognitive abilities, with reported increases across the life span until reaching about .80 in adulthood. When compared with estimates for specific cognitive abilities (ranging from .33 to .68), these findings were interpreted to suggest that the nature of genetic

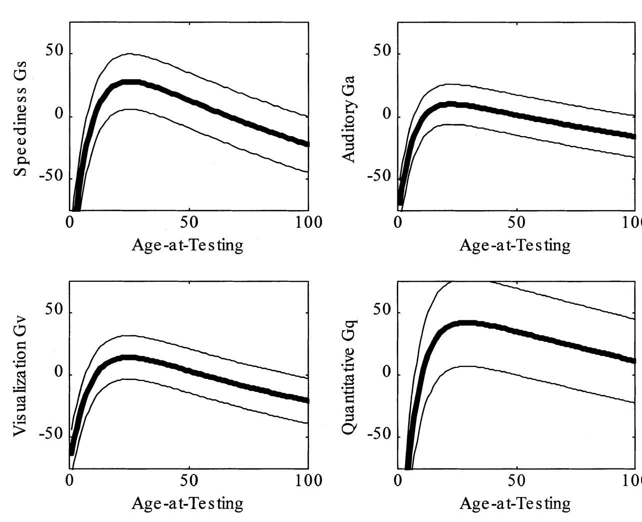

Figure 11. Expected latent curves for *Gs*, *Ga*, *Gv*, and *Gq*.

influence in cognitive abilities is more general than specific. Similarly, many of the genetic effects on specific abilities appear to be explained by genetic effects on *g,* although independent genetic influences also exist for specific abilities (Pedersen, Plomin, & McClearn, 1994). In contrast, when we examine longitudinal data on several broad cognitive abilities, the models used here suggest that the biometric components do not exhibit the same developmental properties (as in McArdle et al., 1998).

The parameters (and related derivations) of these kinds of nonlinear models turn out to be useful indices in the description of a life span growth curve. The resulting nonlinear curve shapes (Figures 8, 9, 10, and 11) are nearly identical to the family of growth curves from the early cross-sectional research of Jones and Conrad (1933; see their Figures 1 through 6). A similar growth curve was formalized by von Bertalanffy (1938) and used to describe changes in weight as a function of competing forces of anabolism and catabolism. A similar dual exponential (but inverted) model was used by Cerella, Rybash, Hoyer, and Commons (1993) to describe the rise and fall (see Horn, 1970, 1972) of information-processing-speed differences over age—the differences between older and younger adults are the net result of accumulated growth and accumulated decrements (also see Kearsley, Buss, & Royce, 1977). A similar model (without initial levels and variance estimates) has also been promoted by Simonton (1984, 1989, 1997) in his studies of career productivity, and the basic concepts used here are much the same.

In this sense, the key contributions of this study come from our use of combined longitudinal and cross-sectional data, our estimation of the variances of the latent changes, our use of formal alternative models, and the statistical evaluation of goodness of fit. The methodological results of this study show that it is practical to use two-occasion accelerated time-lag longitudinal data with standard multilevel software to estimate latent growth models of considerable complexity. We started with models based on polynomial age changes, moved to models with linear age segments (i.e., splines), and found a nonlinear model (i.e., a dual exponential model) to describe the growth and decline of multiple factors. The multilevel model comparisons of curves show that these cognitive functions are separable entities that have different growth patterns. At the very least, these growth curve parameters provide some normative information about the dynamics within each of several measures of cognitive functioning. Because of the unusually wide spread of initial ages, a great deal of cross-sectional (between-persons) information was available here. The parameters estimated from these two-occasion data represent one approach toward an accurate representation of the longitudinal (within-person) information as well. As such, these growth and decline curves represent as much precision as possible until more longitudinal data are collected.

Limitations of the Results

The fitting of longitudinal models when data are incomplete is a complex problem, and the test–retest analyses presented here rely on recent statistical developments. Solutions for dealing with incomplete data have been stated in different ways in different disciplines—for example, the Pearson, Aiken, Lawley, and Meredith selection theorems and Rubin's theorems for incomplete groups (for overviews, see Little & Rubin, 1987; McArdle, 1994; McArdle & Cattell, 1994). In practice, we assume (a) that the initial selection into the study (i.e., before the first occasion of measurement) is an adequate sampling from the population of interest, (b) that this population has the same growth model for all incomplete subgroups ($g = 1$ to G), and (c) that we can restrict the age basis of the growth pattern (e.g., using the previous models). These modern statistical approaches allow opportunities for use of all the available data and address some problems of subject attrition and refusals (Brown, Indurkhya, & Kellam, 2000; Little & Rubin, 1987; McArdle & Hamagami, 1991, 1992, 1996). However, all of the models here are based on these untestable assumptions.

The results presented here also rely on the contemporary developments in item response scaling offered by the WJ–R scales, and this reliance would be a limitation when fitting latent growth models to other scales. That is, because these are single-dimensional and interval-scaled tests (i.e., Rasch scales), we are more confident in mapping the magnitude of change from one score to another onto the same ability metric even though we are using different sections of the scale (i.e., some different items) and different variables (i.e., all different items). Most longitudinal models presume these conditions of measurement, but few longitudinal data sets meet these stringent assumptions. On the other hand, the WJ–R has not yet been widely used in longitudinal analyses, which means that the description of the growth curves, especially the growth and decline parameters, may not directly translate onto other more traditional measures. Although we think it is likely that the linkage of the WJ–R to other measures is practical (e.g., Aggen, 1998; Flanagan & McGrew, 1998; McArdle & Woodcock, 1997; Shaywitz et al., 1992; Woodcock, 1990), it is probably not true for all constructs of interest to cognitive researchers.

These multilevel longitudinal growth curve results also rely on our assumption of the metric factorial invariance (Horn & McArdle, 1992; McArdle & Cattell, 1994; Meredith, 1993) of the WJ–R tests over time and age. In all growth analyses presented here, we assumed the WJ–R scale measured the same construct at all ages in the life span—from ages 2 to 95—and no model was presented here for checking these critical assumptions. Here we relied on previous analyses of the larger set of data presented by McGrew et al. (1991) and suggested reasonable levels of invariance across multiple age groups (Grades K–3, 4–7, and 8–12, and ages 18–40, 41–59, and 60+ years) for both 16 tests and 27 tests (pp. 163–179). A more detailed analysis of factor invariance over time suggests that invariance can be achieved at almost any time lag as long as some practice components are included (McArdle & Nesselroade, in press).

Another threat to the validity of these interpretations of age growth and decline patterns is that the time lag in this longitudinal study may not have been long enough (<10 years) to eliminate all practice confounds or to examine secular or cohort changes. Of special importance here are the age-sensitive determination of cohort changes (e.g., Boyne, 1960; Nesselroade & Baltes, 1979; Schaie, 1996) and some evaluation of what is now known as the *Flynn effect*—large increases in average intellectual abilities over successive age cohorts throughout the 20th century (Flynn, 1999; Neisser, 1998). There is some doubt about the overall magnitude of these effects, the intellectual ability constructs that are changing the most, and the persons who are affected. However, our current longitudinal study was not long enough in time to include cohort parameters and to quantify these hypotheses (e.g., McArdle & Anderson, 1990; Miyazaki & Raudenbush, 2000). Because of our use of aging persons, the possibility of subject mortality and "terminal drop" is a potential problem as well (see Bosworth, Schaie, & Willis, 1999; Maier & Smith, 1999; Siegler & Botwinick, 1979).

More specific questions about the likely heterogeneity of subjects were not examined here, and these could have profound impacts on the growth curves. Parameters can be added to these multilevel longitudinal models to compare growth and changes associated with gender, ethnic differences, geographical areas, and educational levels. But other observed groupings may account for the variation in the curve parameters, including family income levels, social networks, health evaluations, and other related information. These models may need different parameters for the means and variances (amplitude and phase differences) or even group difference in the rates (and, hence, peak ages). New forms of mixture distribution analysis may be required to detect latent group differences in trajectories and dynamics to quantify a set of optimal classification profiles from available demographic variables (L. K. Muthén & Muthén, 1998; Nagin, 1999).

Future Research

The accelerated two-occasion data used here provide the initial basis for the measurement of developmental change even when additional longitudinal measurements are obtained (Burr & Nesselroade, 1990). Of course, choosing the most informative interval of time between these tests is a complex theoretical problem that is not the same for all measures or ages (see Cattell, 1957; McArdle & Woodcock, 1997). In these data, as usual, the relationships between the cognitive factors and other achievement clusters 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 model can be used to determine the minimum aggregation of time lag needed to accurately estimate any change patterns (as in McArdle, 1994; McArdle & Hamagami, 1996).

Previous research suggests that many different forms of incomplete data models can have reasonable power (Brown et al., 2000; McArdle & Hamagami, 1992; B. O. Muthén & Curran, 1997), and this was illustrated here. The resulting power to test basic growth hypotheses varies as a function of (a) the test of means and variances, (b) the type of time-lag pattern selected, (c) the number of occasions of measurement, and (d) the communality of the variables used to indicate the common factors. Previous researchers have pointed out both problems of time-lag designs (e.g., Helms, 1992; Overall, 1987; Schlesselman, 1973), but it seems reasonable to use the time-lag design when relatively little is known about the characteristics of the tests or the traits.

More elaborate models of developmental change may be needed to account for other important features of tests and traits. As more occasions of measurement become available, more complex nonlinear models can be formulated and evaluated (e.g., Nesselroade & Boker, 1994). From a methodological perspective, the nonlinear mixed models used here are also directly related to more general and complex nonlinear models for lengthy longitudinal series, such as the coupled oscillator model by Nesselroade and Boker (1994; with one cycle) or the one-compartment fluid intake–elimination pharmokinetics model of Pinherio and Bates (2000). In addition, there exists a more general family of growth functions offered by Richards (1959), Preece and Baines (1978), Sandland and McGilchrist (1979), and Bock (1989, 1991; Bock et al., 1973). Each of these prior models can be represented in this same multilevel statistical fashion, and each may apply here.

Of more general theoretical importance in this research is a related but more complex set of questions: How do these functions relate to one another (longitudinally) within an individual? and What are the time-dependent dynamic sources of variation in these changes? By fitting these dynamic models we can further examine Cattell's (1971) investment theory, which suggests that *Gf* is the leader or key source of the development of other cognitive functions. Evidence pertaining to the cross-variable dynamic predictions of *Gf-Gc* theory is largely lacking, but some studies have focused on examining these hypotheses. The recent study by Sliwinski and Buschke (1999) examined a particular kind of dynamic hypothesis from the point of view of a processing speed hypothesis (Salthouse, 1996), using multilevel models of time with initial age and processing speed as predictor variables. Although these authors did not directly address the *Gf-Gc* question, the multilevel methodology used here can be extended to permit such an investigation (e.g., McArdle, 2001; McArdle & Hamagami, 2001; McArdle et al., 2000; McArdle & Nesselroade, 1994). These newer models are not simple applications of multilevel models but attempt to untangle some of these issues and make a clearer separation of normative and nonnormative age changes.

The scientific evaluation of these kinds of growth and change questions can require many different approaches for the analysis of longitudinal multivariate data. We think that many uses of accelerated longitudinal data collection design (Bell, 1954) can play an important role in this future research, and out of necessity, the benefits begin with a second occasion of measurement. We hope these structural analyses of growth and change using WJ–R information can be substantively useful.

References

Aber, M. S., & 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.

Aggen, S. H. (1998). Interbattery convergent measurement models applied to the WAIS and WJ–R. In J. J. McArdle & R. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 197–214). Mahwah, NJ: Erlbaum.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. In C. C. Clogg (Ed.), *Sociological methodology, 1990* (pp. 93–114). San Francisco: Jossey-Bass.

Anstey, K. J., Luszcz, M. A., & Sanchez, L. (2001). A reevaluation of the common factor theory of shared variance among age, sensory function, and cognitive function in older adults. *Journal of Gerontology: Psychological Sciences, 56,* 3–11.

Baltes, P. B., & Mayer, K. U. (Eds.). (1999). *The Berlin aging study: Aging from 70 to 100.* New York: Cambridge University Press.

Baltes, P. B., & Schaie, K. W. (1976). On the plasticity of intelligence in adulthood and old age: Where Horn and Donaldson fail. *American Psychologist, 31,* 720–725.

Bayley, N. (1966). Learning in adulthood: The role of intelligence. In H. J. Klausmeier & C. W. Harris (Eds.), *Analyses of concept learning.* New York: Academic Press.

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

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

Bock, R. D. (Ed.). (1989). *Multilevel analysis of educational data.* San Diego, CA: Academic Press.

Bock, R. D. (1991). Prediction of growth. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 126–136). Washington, DC: American Psychological Association.

Bock, R. D., Wainer, H., Peterson, A., Thissen, D., Murray, J., & Roche, A. (1973). A parameterization for individual human growth curves. *Human Biology, 45,* 63–80.

Boker, S. M., & McArdle, J. J. (1995). A statistical vector field analysis of longitudinal aging data. *Experimental Aging Research, 21,* 77–93.

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

Botwinick, J. (1977). Intellectual abilities. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 580–605). New York: Van Nostrand Reinhold.

Boyne, A. W. (1960). Secular changes in the stature of adults. In J. M. Tanner (Ed.), *Human growth: Symposia of the Society for the Study of Human Biology* (Vol. 3, pp. 97–120). New York: Pergamon Press.

Brown, C. H., Indurkhya, A., & Kellam, S. G. (2000). Power calculations for data missing by design: Applications to a follow-up study of lead exposure and attention. *Journal of the American Statistical Association, 95,* 383–395.

Browne, M. W., & Du Toit, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. 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. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147–158.

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

Burr, J. A., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), *New statistical methods in developmental research* (pp. 3–34). New York: Academic Press.

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

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–24). Mahwah, NJ: Erlbaum.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. [Abstract] *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: World.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action.* Amsterdam: North-Holland.

Cattell, R. B. (1998). Where is intelligence? Some answers form the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 29–38). Mahwah, NJ: Erlbaum.

Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life-span. *Acta Psychologica, 86,* 109–197.

Cerella, J., Rybash, J. M., Hoyer, W., & Commons, M. L. (Eds.). (1993). *Adult information processing: Limits on loss.* San Diego, CA: Academic Press.

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

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Collins, L. M., & Horn, J. L. (Eds.). (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions.* Washington, DC: American Psychological Association.

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

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research, 31,* 371–403.

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature, 388,* 468–471.

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

Draper, N., & Smith, H. (1981). *Applied regression analysis.* New York: Wiley.

Duncan, S. C., & Duncan, T. E. (1995). Modeling the processes of development via latent variable growth curve methodology. *Structural Equation Modeling: A Multidisciplinary Journal, 2,* 187–213.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: Erlbaum.

Eichorn, D. H., Clausen, J. A., Haan, N., Honzik, M. P., & Mussen, P. (Eds.). (1981). *Present and past in middle life.* New York: Academic Press.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 4,* 341–349.

Flanagan, D. P., & McGrew, B. (1996). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314–325). New York: Guilford Press.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc theory: Joint confirmatory factor analysis of the WJ-R and KAIT in a non-White sample. *Journal of School Psychology, 36,* 151–182.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54,* 5–20.

Giambra, L. M., Arenberg, D., Zonderman, A. B., Kawas, C., & Costa, P. T. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychology and Aging, 10,* 123–139.

Goldstein, H. (1995). *Multilevel statistical models.* New York: Oxford University Press.

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

Helms, R. (1992). Intentionally incomplete longitudinal designs: I. Methodology and comparison of some full span designs. *Statistics in Medicine, 2,* 1889–1913.

Horn, J. L. (1970). Organization of data on life-span development of human abilities. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology*. New York: Academic Press.

Horn, J. L. (1971). Integration of concepts of reliability and standard error of measurement. *Educational & Psychological Measurement, 31,* 57–74.

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

Horn, J. L. (1985). Remodeling old models of intelligence: Gf-Gc theory. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.

Horn, J. L. (1986). Intellectual ability concepts. In R. L. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 25–77). Hillsdale, NJ: Erlbaum.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade (Ed.), *Handbook of multivariate psychology* (pp. 645–685). New York: Academic Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *Woodcock–Johnson technical manual* (pp. 197–246). Allen, TX: DLM Teaching Resources.

Horn, J. L. (1998). A basis for research on age differences in cognitive abilities. In J. J. McArdle & R. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57–92). Mahwah, NJ: Erlbaum.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253–270.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107–129.

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 1980s: Psychological issues* (pp. 503–541). Washington, DC: American Psychological Association.

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

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hox, J. J., & Kreft, I. G. G. (1994). Multilevel analysis methods introduction: Special issue on multilevel analysis. *Sociological Methods and Research, 22,* 283–300.

Humphreys, L. G. (1989). Intelligence: Three kinds of instability and their consequences for policy. In R. Linn (Ed.), *Intelligence: Measurement, theory and public policy* (pp. 193–216). Urbana: University of Illinois Press.

Hunt, E. B. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce.* New York: Russell Sage Foundation.

Hunt, W. L. (1949). The relative rates of decline of Wechsler-Bellevue "hold" and "don't hold" tests. *Journal of Consulting Psychology, 13,* 440–443.

Jensen, A. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jones, H. E., & Conrad, H. S. (1933). The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychology Monographs, 13*(1), 223–275.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Scientific Software International.

Kangas, J., & Bradway, K. P. (1971). Intelligence at middle age: A thirty-eight year follow-up. *Developmental Psychology, 5,* 333–337.

Kearsley, G. P., Buss, A. R., & Royce, J. R. (1977). Developmental change and the multidimensional cognitive system. *Intelligence, 1,* 257–273.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling (introducing statistical methods).* London: Sage Publications.

Kreft, I., De Leeuw, J., & van der Leeden, R. (1994). Review of five analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *American Statistician, 48,* 324–335.

Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in the old and very old: Cross-sectional results from the Berlin Aging Study. *Psychology and Aging, 12,* 410–432.

Lindsey, J. K. (1993). *Models for repeated measures.* Oxford, England: Clarendon Press.

Littell, R. C., Miliken, G. A., Stoup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

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

Maier, H., & Smith, J. (1999). Psychological predictors of mortality in old age. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences, 54B,* 44–54.

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

McArdle, J. J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research, 25,* 81–87.

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

McArdle, J. J. (1998). Contemporary statistical models of test bias. In J. J. McArdle & R. W. Woodcock (Eds.), *Human abilities in theory and practice* (pp. 157–195). Mahwah, NJ: Erlbaum.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 341–380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J. (in press). Growth curve analysis. In N. J. Smelser & P. B. Baltes (Eds.), *The international encyclopedia of the social & behavioral sciences.* New York: Pergamon Press.

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

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

McArdle, J. J., & Bell, R. Q. (2000). Recent trends in modeling longitudinal data by latent growth curve methods. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and scientific examples* (pp. 69–108). Mahwah, NJ: Erlbaum.

McArdle, J. J., & Cattell, R. B. (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, E. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 276–304). Washington, DC: American Psychological Association.

McArdle, J. J., & Hamagami, E. (1992). Modeling incomplete longitudinal and cross-sectional 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., & Hamagami, F. (2001). Linear dynamic analyses of incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *Methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.

McArdle, J. J., Hamagami, F., Elias, M. F., & Robbins, M. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research, 17,* 29–52.

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

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., & Nesselroade, J. R. (in press). Growth curve analysis in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology: Vol. 3, Research methods in psychology*. New York: Pergamon Press.

McArdle, J. J., Paskus, T. A., & Boker, S. M. (in press). A multilevel multivariate analysis of academic success in college based on NCAA student-athletes. *Multivariate Behavioral Research.*

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

McArdle, J. J., Prescott, C. A., Hamagami, F., & Horn, J. L. (1998). A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology, 14,* 69–114.

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

McArdle, J. J., & Woodcock, R. W. (Eds). (1998). *Human cognitive abilities in theory and practice*. Mahwah, NJ: Erlbaum.

McArdle, J. J., & Woodcock, R. W. (2000). *Longitudinal multilevel analyses of test–retest data on the growth and decline of cognitive abilities*. Manuscript submitted for publication.

McCrae, R. R., Arenberg, D., & Costa, P. T. (1987). Declines in divergent thinking with age: Cross-sectional, longitudinal, and cross-sequential analyses. *Psychology and Aging, 2,* 130–137.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston: Allyn & Bacon.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock–Johnson technical manual*. Allen, TX: DLM Teaching Resources.

Menard, J. (2000). Coefficients of determination for multiple logistic regression analysis. *American Statistician, 54,* 17–24.

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

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

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

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

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

Muthén, L. K., & Muthén, B. O. (1998). *Mplus, the comprehensive modelling program for applied researchers: User's guide.* Los Angeles, CA: Muthén & Muthén.

Nagin, D. (1999). Cluster analysis of longitudinal curve components. *Psychological Methods, 4,* 139–157.

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

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

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

Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 121–147). Washington, DC: American Psychological Association.

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

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

Overall, J. E. (1987). Estimating sample size for longitudinal studies of age-related cognitive decline. *Journal of Gerontology, 42*(2),137–141.

Pedersen, N. L., Plomin, R., & McClearn, G. E. (1994). Is there G beyond g? (Is there genetic influence on specific cognitive abilities independent of genetic influence on general cognitive ability?). *Intelligence, 18,* 133–143.

Pinherio, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* New York: Springer.

Plomin, R., Pedersen, N. L., Lichtenstein, P., & McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behavior Genetics, 24,* 207–215.

Pothoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis model useful especially for growth curve problems. *Biometrics, 51,* 313–326.

Preece, M. A., & Baines, M. J. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology, 5,* 1–24.

Rao, C. R. (1958). Some statistical methods for the comparison of growth curves. *Biometrics, 14,* 1–17.

Raudenbush, S. W., & Chan, W. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology, 61,* 941–951.

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

Rencher, A. C. (1995). *Methods of multivariate analysis.* New York: Wiley.

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany, 10,* 290–300.

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

Sandland, R. L., & McGilchrist, C. A. (1979). Stochastic growth curve analysis. *Biometrics, 35,* 255–271.

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

Schlesselman, J. J. (1973). Planning a longitudinal study: 2. Frequency of measurement and study duration. *Journal of Chronic Disease, 26,* 561–570.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., &

Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326,* 145–150.

Siegler, I. C., & Botwinick, J. (1979). A long-term longitudinal study of intellectual ability of older adults: The matter of selective subject attrition. *Journal of Gerontology, 34,* 242–245.

Simonton, D. K. (1984). Creative productivity and age: A mathematical model based on a two-step cognitive process. *Developmental Review, 4,* 77–116.

Simonton, D. K. (1989). Age and creative productivity: Nonlinear estimation of an information processing model. *International Journal of Aging and Human Development, 29,* 23–37.

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review, 104,* 66–89.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24,* 323–355.

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

Smith, P. L. (1979). Splines as a useful and convenient tool. *The American Statistician, 32,* 57–63.

Snijders, T., & Bosker, R. (1994). Modeled variance in two-level models. *Sociological Methods and Research, 22,* 342–363.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Stimson, J. A., Carmines, E. G., & Zeller, R. A. (1978). Interpreting polynomial regression. *Sociological Methods and Research, 6,* 515–524.

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

Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental Psychology, 35,* 986–1000.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

von Bertalanffy, L. (1938). A quantitative theory of organic growth. *Human Biology, 10,* 181–243.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: The Psychological Corporation.

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

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116,* 363–381.

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement, 20,* 59–69.

Wilson, R. S., Gilley, D. W., Bennett, D. A., Beckett, L. A., & Evans, D. A. (2000). Person-specific paths of cognitive disease in Alzheimer's disease and their relation to age. *Psychology and Aging, 5,* 18–28.

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

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 Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–127). Mahwah, NJ: Erlbaum.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery— Revised.* Allen, TX: DLM Teaching Resources.

Zeger, S. L., & Harlow, S. D. (1987). Mathematical models from laws of growth to tools for biological analysis. *Biometrika, 30,* 16–28.

Zuker, D. M., Zerbe, G. O., & Wu, M. C. (1995). Inference for the association between coefficients in a multivariate growth curve model. *Biometrics, 51,* 413–424.

Appendix A

Computer Programming for Nonlinear Mixed Models

The SAS PROC MIXED computer program was used to estimate the parameters and goodness of fit. Excellent examples of the use of these programs can be found in Singer (1998) and Littell, Miliken, Stoup, and Wolfinger (1996). A brief example of the SAS code we used follows:

```
TITLE1 'Retest Study Polynomial Models';
DATA tmp1;
        SET ngcs.retest_2000;
        agec1 = (tage1 - 19); prac1=0;
        sega1 = agec1; IF (agec1 GT 0) THEN sega1 = 0;
        segb1 = agec1; IF (agec1 LT 0) THEN segb1 = 0;
        agec2 = (tage2 - 19); prac2=1;
        sega2 = agec2; IF (agec2 GT 0) THEN sega2 = 0;
        segb2 = agec2; IF (agec2 LT 0) THEN segb2 = 0;
        FILE outfile LRECL=200 LINESIZE=200;
        PUT
        #1 id agec1 prac1 sega1 segb1 bca_w1
        #2 id agec2 prac2 sega2 segb2 bca_w2;
        RUN;
DATA tmp3;
        INFILE outfile LRECL=200 LINESIZE=200;
        INPUT agec01 practice sega segb y01;
```

```
        RUN;
TITLE2 'Model 1; Linear Age+Corr';
PROC MIXED NOCLPRINT METHOD=ML COVTEST IC; CLASS id;
        MODEL y01 = agec01 / SOLUTION DDFM=BW CHISQ;
        RANDOM INTERCEPT agec01 / SUBJECT=id TYPE=UNR;
        RUN;
TITLE2 'Model 2: Quadratic Age+ ';
PROC MIXED NOCLPRINT METHOD=ML COVTEST IC; CLASS id;
        MODEL y01 = agec01 agec01*agec01 / SOLUTION DDFM=BW CHISQ;
        RANDOM INTERCEPT agec01 / SUBJECT=id TYPE= UNR;
        RUN;
TITLE2 'Model 3: Dual Segmented Age + Restricted Corrs';
PROC MIXED NOCLPRINT METHOD=ML COVTEST IC; CLASS id;
        MODEL y01 = sega segb / SOLUTION DDFM=BW CHISQ;
        RANDOM INTERCEPT sega segb / SUBJECT=id TYPE= UN GCORR;
        PARMS (260) (1) (2) (1) (0) (2) (40) / EQCONS=5;RUN;
TITLE2 'Model 4: Dual Segmented Age+Practice+Corr Restricted ';
PROC MIXED NOCLPRINT METHOD=ML COVTEST IC; CLASS id;
        MODEL y01 = sega segb practice / SOLUTION DDFM=BW CHISQ;
        RANDOM INTERCEPT sega segb practice / SUBJECT=id TYPE=UN GCORR;
        PARMS (68) (1) (2) (1) (0) (2) (0) (0) (5) (75) / EQCONS=5,7,8,9;  RUN;
```

The newer nonlinear version of this program, SAS PROC NLMIXED, has only recently been available, so previous applications are sparse. However, this program was required for the estimation of parameters and goodness of fit in the dual exponential model, and an example of the code we used follows:

```
TITLE2 'Model 5: ExpExp Nonlinear Model with Individual Differences ';
PROC NLMIXED;
       traject = level + slope (EXP(-rate_g*tage01) - EXP(-rate_d*tage01));
       MODEL y01 ~ NORMAL(traject, v_error);
       RANDOM level slope ~ NORMAL([m_level, m_slope],
             [v_level, c_levslo, v_slope]) SUBJECT=id;
        PARMS m_level =-80 m_slope = 120 rate_g =.001 rate_d =.100
             v_error=20 v_level=80 v_slope=10 c_levslo=-.01; RUN;

TITLE2 'Model 6: ExpExp Nonlinear Model with Practice Mean ';
       PROC NLMIXED;
       traject = level + slope (EXP(-rate_g*tage01) - EXP(-rate_d*tage01)) +
(m_prac*practice);

MODEL y01 ~ NORMAL (traject, v_error);
RANDOM level slope ~ NORMAL([m_level, m_slope],
       [v_level, c_levslo, v_slope]) SUBJECT=id;
  PARMS m_level =-80 m_slope = 120 rate_g =.001 rate_d = .100 m_prac=.01
       v_error=20 v_level=80 v_slope=10 c_levslo= -.01; RUN;
```

For further examples and details of alternative codes, see Pinherio and Bates (2000), and see the other SAS codes included on our website: http://kiptron.psyc.virginia.edu.

Appendix B

Numerical Results From Linear and Nonlinear Models

Numerical results from multilevel polynomials (not listed in tabular form here). The resulting estimates for the fourth-order growth curve for the BCA variable can be expressed as

$$\underset{[\pm\sigma_{BCA[m]}]}{\mu_{BCA[m]}} = \underset{[\pm 8.0]}{17.9} + \underset{[\pm 0.16]}{(Age_m \cdot 1.17)} + (Age_m^2 \cdot -0.100) + (Age_m^3 \cdot 0.0023) + (Age_m^4 \cdot -0.00002|) + \underset{[\pm 4.1]}{e_m},$$

where the estimated latent mean values are listed and the estimated latent standard deviations are included below them in brackets (i.e., [±]). In contrast, the estimated fourth-order curves for the *Gf* and *Gc* variables can be written as

(*Appendix continues*)

$$\underset{[\pm\sigma_{Gf[m]}]}{\mu_{Gf[m]}} = \underset{[\pm 11.8]}{18.2} + \underset{[\pm 0.21]}{(Age_m \cdot 1.15)} + (Age_m^2 \cdot -0.114) + (Age_m^3 \cdot 0.0026) + (Age_m^4 \cdot -0.00002) + \underset{[\pm 8.7]}{e_m}$$

and

$$\underset{[\pm\sigma_{Gc[m]}]}{\mu_{Gc[m]}} = \underset{[\pm 14.3]}{34.1} + \underset{[\pm 0.32]}{(Age_m \cdot 2.43)} + (Age_m^2 \cdot -0.153) + (Age_m^3 \cdot 0.0032) + (Age_m^4 \cdot -0.00002) + \underset{[\pm 7.1]}{e_m} .$$

Numerical results from age-spline models (see Table 7 and Figure 7B). The final bilinear equation for the BCA scores is written as

$$\underset{[\pm\sigma_{\mathrm{BCA}[m]}]}{\mu_{\mathrm{BCA}[m]}} = \underset{[\pm 8.4]}{20.6} + \underset{[\pm 1.16]}{(\{Age_m - 20 < 0\} \cdot 2.64)} + \underset{[\pm 0.22]}{(\{Age_m - 20 > 0\} \cdot -0.39)} + (p[0] \cdot 3.2) + \underset{[\pm 6.0]}{e_m} .$$

The estimated bilinear equations for *Gf* and *Gc* are

$$\underset{[\pm\sigma_{Gf[m]}]}{\mu_{Gf[m]}} = \underset{[\pm 9.9]}{16.9} + \underset{[\pm 1.40]}{(\{Age_m - 20 < 0\} \cdot 2.68)} + \underset{[\pm 0.29]}{(\{Age_m - 20 > 0\} \cdot -0.45)} + (p[0] \cdot 6.2) + \underset{[\pm 6.8]}{e_m}$$

and

$$\underset{[\pm\sigma_{Gc[m]}]}{\mu_{Gc[m]}} = \underset{[\pm 18.2]}{41.4} + \underset{[\pm 1.87]}{(\{Age_m - 20 < 0\} \cdot 5.11)} + \underset{[\pm 0.55]}{(\{Age_m - 20 > 0\} \cdot -0.08)} + (p[0] \cdot 1.8) + \underset{[\pm 6.8]}{e_m} .$$

Numerical results from nonlinear models (see Table 8 and Figure 7D). The final dual exponential model estimates for BCA can be written as

$$\underset{[\pm\sigma_{\mathrm{BCA}[m]}]}{\mu_{\mathrm{BCA}[m]}} = \underset{[\pm 9.6]}{-75.2} + \{[\exp(-.0066 \cdot Age_m) - \exp(-.1165 \cdot Age_m)]\} \underset{[\pm 0.79]}{117.7} + (p[0] \cdot 2.7) + \underset{[\pm 3.6]}{e_m} .$$

The final nonlinear model estimates for *Gf* and *Gc* can be written as

$$\underset{[\pm\sigma_{Gf[m]}]}{\mu_{Gf[m]}} = \underset{[\pm 10.5]}{-116.5} + \{[\exp(-.0052 \cdot Age_m) - \exp(-.1539 \cdot Age_m)]\} \underset{[\pm 4.1]}{156.3} + (p[0] \cdot 3.0) + \underset{[\pm 8.2]}{e_m}$$

and

$$\underset{[\pm\sigma_{Gc[m]}]}{\mu_{Gc[m]}} = \underset{[\pm 17.1]}{-116.0} + \{[\exp(-.0026 \cdot Age_m) - \exp(-.1104 \cdot Age_m)]\} \underset{[\pm 2.20]}{179.2} + (p[0] \cdot 0.0) + \underset{[\pm 7.0]}{e_m} .$$

Numerical results for common factor growth function (not in tabular form; see Footnote 5 in the text). The final common factor model estimates can be written as

$$\underset{[\pm\sigma_{g[m]}]}{\mu_{g[m]}} = \{[\exp(-.0053 \cdot Age_m) - \exp(-.1085 \cdot Age_m)]\} \underset{[\pm 0.79]}{117.7} + \underset{[\pm 3.6]}{u_m}$$

and

$$\underset{[\pm\sigma_{Gf[m]}]}{\mu_{Gf[m]}} = \underset{[\pm 10.5]}{-116.5} + 1(\mu_{g[m]}) + \underset{[\pm 4.1]}{3.0} + \underset{[\pm 8.2]}{u_{Gf[m]}}$$

and

$$\underset{[\pm\sigma_{Gc[m]}]}{\mu_{Gc[m]}} = \underset{[\pm 10.5]}{-116.5} + \underset{[\pm 4.1]}{1.55\,(\mu_{g[m]})} + \underset{[\pm 8.2]}{3.0} + u_{Gc[m]} .$$

Received December 5, 2000
Revision received August 3, 2001
Accepted August 20, 2001 ■