Psychology and Aging
2004, Vol. 19, No. 2, 243–259

0882-7974/04/$12.00 DOI: 10.1037/0882-7974.19.2.243

Modeling Age and Retest Processes in Longitudinal Studies of Cognitive Abilities

Emilio Ferrer
University of California, Davis

Timothy A. Salthouse
University of Virginia

Walter F. Stewart and Brian S. Schwartz
Johns Hopkins University

Mixed effects models were used to examine the separate effects associated with age and retest on changes in various cognitive abilities. Individuals ($N > 800$) ranging in age from 40 to 70 years at the 1st measurement occasion were assessed with measures of memory, spatial abilities, and speed on 4 occasions. All cognitive abilities showed decline associated with increased age and improvement across the 4 measurement occasions. The age-related effects were similar across variables, but the practice effects were largest for memory and smallest for speed. When retest effects were not included in the models, the age-related effects were underestimated, with the magnitude of bias depending on the size of the ignored retest effects. It is suggested that both age and retest should be modeled simultaneously when analyzing longitudinal data because part of the change across occasions may be attributable to practice or reactive effects.

The advocacy of longitudinal, versus cross-sectional, approaches has a long history in the study of cognitive abilities (e.g., Nesselroade & Baltes, 1979). There now seems to exist a general agreement on the benefits of longitudinal designs for examining both growth and decline aspects of change. Some of these benefits include the possibility of estimating within-person change, individual differences in changes, and correlations between rates of change and other variables (see, e.g., Schaie, 1983; Schaie & Hofer, 2001), and the possibility of integrating individual differences within developmental change (Wohlwill, 1970). Despite these clear advantages, however, longitudinal studies also bring with them a new set of methodological issues. Among the most important of these are attrition, cohort effects, and practice or retest effects (Cohen, 1991; Nesselroade & Baltes, 1979). In this article, we focus on practice—or retest—effects in longitudinal studies. In particular, we illustrate how recent analytical procedures can be used to separate the effects of age and retest on changes in various cognitive abilities.

Emilio Ferrer, Department of Psychology, University of California, Davis; Timothy A. Salthouse, Department of Psychology, University of Virginia; Walter F. Stewart, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health; Brian S. Schwartz, Division of Occupational and Environmental Health, Johns Hopkins Bloomberg School of Public Health.

Support in the preparation of this article was provided by Grants T32 AG20500–01 and AG10785 from the National Institute on Aging.

We thank Aki Hamagami and Jack McArdle for their comments on drafts of this article.

Correspondence concerning this article should be addressed to Emilio Ferrer, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616. E-mail: eferrer@ucdavis.edu

Practice Effects

Researchers studying the development of a particular attribute collect data on individuals at different occasions. The repeated observations permit examination of changes in the persons' attribute over time. In some instances, the assessment does not interfere with the attribute itself. If one considers, say, physical growth, a person will grow independent of the number of assessments—or of the method—used to measure growth. In other instances, however, because of the technique, the number of assessments, or both, the measurement may interfere with the developmental process at hand. This is likely to be the case with cognitive abilities. At each assessment, the person is presented with testing tasks, which might also serve as learning materials. It is therefore possible that at the next occasion of measurement the person's performance will be influenced by the previous experience. That is, the previous exposure to the testing materials may be helping the person to improve and, thus, interfering with the normal development the researcher is trying to capture. The extent of this contamination will depend on various factors, including the nature of the attribute being measured, the number of repeated assessments and the retest interval between them, and the kind of assessment (see, e.g., Cattell, 1957).

The phenomenon of practice, or retest, effects is well-known in the area of cognitive abilities (e.g., Horn, 1972; McArdle & Woodcock, 1997; Nesselroade & Baltes, 1979; Schaie, 1996; Thorndike, 1933). It is common to find that participants' performance in a cognitive test improves over occasions, with differences in the magnitude of improvement across variables (e.g., Lowe & Rabbitt, 1998; Rabbitt, 1993; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001; Zelinski, Gilewski, & Schaie, 1993) and, presumably, across persons. For example, both the means and the variances have been found to increase from the first to the second testing occasion (Jones, 1962; McArdle & Woodcock, 1997), and practice effects have been detected over four occasions, although they decreased in magnitude across occasions (Rabbitt et al., 2001). If one is interested in the development of cognitive abilities, therefore, the changes that are due to a person's increase in age (i.e., presumably reflecting development or maturation) need to be separated from the changes that are due to retesting (i.e., presumably reflecting practice and experience).

Separating Age Effects From Retest Effects

The attempt to separate age effects from retest effects has a history parallel to the study of the development of cognitive abilities (see, e.g., Donaldson & Horn, 1992). One approach to separating the two effects has tried to avoid practice effects by designing studies with large time intervals between occasions (Schaie, 1996). The underlying assumption of this approach is that given ample time, any effects associated with practice will dissipate. One problem with this method, however, is the uncertainty with regard to how large the intervals need to be to exclude any practice effect. Retest effects have been found in studies with retest intervals of several years (see Rabbitt et al., 2001), and some researchers have suggested that with certain variables, 6 years may be needed between occasions to eliminate practice effects (Zelinski & Burnight, 1997). Moreover, optimal time intervals are likely to differ across measures, ages, and persons (see Cattell, 1957), and very long intervals may reduce the sensitivity to detect change, especially in periods of accelerated growth or decline.

Separating age changes from practice effects has also been attempted via statistical modeling (e.g., McArdle & Anderson, 1990; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; McArdle & Woodcock, 1997; Rabbitt et al., 2001; Wilson et al., 2002; see also Cattell, 1970). This approach consists of including separate terms for age and occasion in the same analytic model, but a potential problem arises when the age increment and the retest occasion are highly correlated. For example, imagine a study in which individuals of the same age are measured at intervals of exactly 1 year. The age-related changes in a particular variable will then be confounded with possible changes in performance that are due to retest. That is, increments in retest will correlate perfectly with increments in age. When the retest interval presents variation and there is ample spread in age, however, this approach presents merits.

McArdle and Anderson (1990) used this method to examine changes in full-scale Wechsler Adult Intelligence Scale (WAIS) scores for a sample of individuals ranging in age from 60 to 75 years. Each individual was measured four times at intervals of 4 years between occasions. Changes in WAIS scores were analyzed by using a latent growth model that combined basis coefficients for age and measurement occasion. The findings indicated a decline in WAIS scores with increasing age, but this decline was somewhat offset by small retest gains. Because age increment and measurement occasion were partially correlated, however, some of the fitted models (i.e., latent functions for both age and retest) were not identified. McArdle and Anderson recognized that both dimensions had not been separated enough in the design of the study and acknowledged the difficulty of using statistical tools to overcome limitations of the study design.

One possible way to avoid a high correlation between age increment and measurement occasion is to introduce variation in the retest interval (McArdle & Woodcock, 1997; Schlesselman, 1973; Vinsonhaler & Meredith, 1966). McArdle and colleagues used this approach by measuring participants on two occasions with retest intervals ranging from several months to 10 years (McArdle et al., 2002; McArdle & Woodcock, 1997). Such a variation enabled the authors to distinguish practice effects from age-related effects in a study involving individuals with a large range of ages (i.e., 2 to 98 years). Because participants were assessed only twice, however, the modeling of practice effects was limited in several regards. First, the only testable hypothesis for practice was a linear effect. Second, practice was forced to be orthogonal with other terms of the model (i.e., intercept and age slope). Third, the model did not include a term for the Age $\times$ Practice interaction, and thus it was not possible to determine whether there were age differences in practice.

Separating growth effects from practice effects in two-occasion data requires that certain restrictions be imposed in the model. When more than two data points are available per person, however, such restrictions can be relaxed and more complex models can be examined. In a recent report, Rabbitt et al. (2001) illustrated this point using data from a 17-year longitudinal study in which participants had been measured up to four times. Rabbitt et al. were interested in examining the effects of age and practice on changes in fluid ability for a large sample of individuals who ranged in age from 40 to 93 years at the initial testing. To this end, they used a random effects model that included age effects (i.e., linear and quadratic terms) together with practice effects (i.e., step increases between successive occasions on which the subject had been measured). The results from this model indicated that fluid abilities decreased over age, with larger declines among older individuals (i.e., negative quadratic effect), but individuals' performance improved across measurement occasions. Moreover, positive retest effects persisted up to the fourth occasion, although the size of the effects decreased across occasions (i.e., 52%, 26%, and 22% of the total effects across the three retests). Rabbitt et al. further attempted to fit a model that considered variation in the practice term, but the estimates were imprecise and no confidence intervals could be estimated. Nevertheless, these results showed that when individuals are measured over time, retest effects can occur even with long intervals, and the magnitude of those effects may differ as a function of the measurement occasion.

Although including both growth and practice in the analytic model can lead to complexities, omitting practice from the model may lead to spurious inferences. For example, if retest effects exist but they are not modeled simultaneously with age, all of the individual change over time is absorbed by the age component. It is therefore possible that negative changes that are due to age become negligible because they are obscured—canceled out—by hidden positive practice effects. As Rabbitt et al. (2001) noted, this pattern could be the reason why various measures of memory (e.g., episodic, text memory, and word recall) do not show age-related declines in some studies (Elias, Robins, & Elias, 1996; Hultsch, Hertzog, Small, McDonald-Miszczk, & Dixon, 1992; Schaie, 1989). Actual declines may have been obscured by performance improvements that were due to retest.

In the current article, we attempt to separate via statistical modeling the effects of age and practice on the changes in various cognitive abilities. We use different models to test specific hypotheses (e.g., linear, latent) about the relation of age and practice to cognitive abilities. We also include a random effect for practice to represent variation in practice effects across individuals. Because of its conceptual logic, practice is not constrained to be orthogonal with other terms in the model (initial level and age slope). That is, individuals who start at higher levels of cognitive functioning may benefit from retest differently than individuals with lower initial ability levels. Finally, to examine a possible relation of practice to age, we examine the interaction between these two terms. This interaction, not examined in earlier studies, tests the extent to which individuals of different ages benefit differently from retest. Although some of the features included here are novel, we build on previous work by others and simply take advantage of newer and more flexible modeling possibilities. The results generated by these models should prove informative about practice effects and possible differences in such effects across persons, measurement occasions, and cognitive abilities.

Method

Participants

This report is based on data from the Age, Lead Exposure, and Neurobehavioral Decline Study (Walter F. Stewart, principal investigator). Extensive information about the study design, participants, and data is available in previously published studies (e.g., Schwartz et al., 2000; Stewart et al., 1999), and thus information about participants and measures is abridged here. The participants were 834 individuals who took part in a 4-year prospective study to evaluate the effects of lead exposure on changes in cognitive function. All participants were men who were between 40 and 70 years of age at the first assessment. Of the 834 participants, 703 were former employees of a U.S. chemical manufacturing facility that produced tetraethyl and tetramethyl lead, and they had ceased occupational lead exposure for an average of 16 years. The other 131 were unexposed individuals.

Measures

Participants visited a clinical setting for up to four occasions, at which time they completed a comprehensive neurobehavioral battery and were assessed on a number of biological measures, including blood lead and bone lead levels. On the basis of the results of an exploratory factor analysis described below, the variables were grouped into three different cognitive abilities. Specifically, several measures from the Rey Auditory Verbal Learning Test (RAVLT; total score across five recall trials, delayed recall, and recognition test score) were used to represent verbal learning and memory. The Delayed Recall subtest from the Rey–Osterreith Complex Figure Test (RCF), the Block Design subtest of the Wechsler Adult Intelligence Scale—Revised (WAIS–R), and the symbol digit paired associate learning test were selected to assess spatial ability. A choice reaction time task, the Stroop Test (A, B, and C forms), the Trail Making Test Parts A and B, and the Digit Symbol subtest from the WAIS–R were used to measure processing speed. The same versions of the tests were administered on each occasion.

Data Description and Preliminary Analyses

Table 1 presents a description of the participants' ages at all occasions of measurement together with the retest intervals between assessments. In addition to the overall sample, information is presented for different age groups (i.e., younger than 50 years, between 50 and 59 years, and 60 years and older). This table also indicates that, of the initial 827 participants who reported their age, 86.7%, 72.3%, and 66.0% were present at the second, third, and fourth assessment occasions, respectively. These retention patterns seem to be uniform across the different age groups, with the possible exception of the younger individuals at the fourth occasion, who showed a somewhat smaller retention rate. Most of the analyses reported below were conducted with maximum-likelihood estimation procedures that are based on all available data. However, it is important to note that similar results were obtained when the analyses were repeated on the 492 individuals with complete data on all four occasions.

Preliminary analyses were initially conducted to select measures that represented the same construct at all occasions. First, z scores based on the score at the first occasion were calculated for all measures, their purpose being to estimate possible changes relative to the initial scores. A factor analysis with oblique rotation was then performed on the data at each of the four occasions to examine what variables could be represented by the same underlying construct. The results suggested that the variables loaded on three factors representing memory, space-related abilities, and processing

Table 1
Description of Ages (in Years) and Time Retests

Age group	Time 1	Time 2	Time 3	Time 4
$<$ 50 years	45.9 (2.84)	47.1 (2.78)	47.9 (2.73)	49.0 (2.68)
Δ_{age}		.987 (.227)	.885 (.154)	.946 (.184)
N	181	153	130	102
50–59 years	55.0 (2.79)	56.1 (2.80)	56.9 (2.84)	57.8 (2.87)
Δ_{age}		1.04 (.274)	.856 (.129)	.923 (.171)
N	357	319	265	249
$>$ 60 years	64.8 (3.21)	65.9 (3.24)	66.5 (3.26)	67.5 (3.25)
Δ_{age}		1.05 (.252)	.835 (.105)	.924 (.193)
N	289	244	203	194
All	56.5 (7.63)	57.5 (7.52)	58.2 (7.50)	59.6 (7.31)
Δ_{age}		1.03 (.258)	.857 (.128)	.927 (.182)
N	827	717	598	545

Note. Values in parentheses are standard deviations.

Table 2
Descriptive Statistics by Age Group and for the Overall Sample

Cognitive ability and age group	Time 1	Time 2	Time 3	Time 4
Memory				
< 50 years	.330 (.841)	.613 (.773)	.794 (.697)	.820 (.777)
50–59 years	.109 (.811)	.402 (.728)	.535 (.735)	.662 (.758)
> 60 years	−.360 (.872)	−.102 (.852)	.092 (.821)	.203 (.927)
All	−.005 (.881)	.275 (.831)	.442 (.802)	.529 (.860)
Space				
< 50 years	.314 (.743)	.359 (.739)	.568 (.722)	.541 (.786)
50–59 years	.130 (.705)	.130 (.726)	.339 (.683)	.296 (.710)
> 60 years	−.360 (.723)	−.365 (.702)	−.220 (.716)	−.230 (.777)
All	−.001 (.769)	.010 (.773)	.199 (.768)	.154 (.805)
Speed				
< 50 years	.338 (.587)	.410 (.700)	.418 (.509)	.426 (.459)
50–59 years	.163 (.578)	.231 (.629)	.252 (.499)	.180 (.804)
> 60 years	−.447 (.835)	−.361 (.781)	−.396 (.748)	−.489 (.897)
All	−.012 (.754)	.066 (.766)	.068 (.686)	−.013 (.868)

Note. Values in parentheses are standard deviations. All values are based on z scores computed from Time 1.

speed, with a similar pattern across occasions.[1] To examine the equivalence of the models across the four occasions, we conducted invariance tests. The first model of these tests was a nonrestrictive model with all the parameters allowed to vary across measurement occasions. This model yielded a fit that served as a baseline against which to compare subsequent models: $\chi^2(248, N = 2{,}716) = 3{,}100$; root mean square error of approximation (RMSEA) = .050. The next model constrained the loadings to be equal, and this restriction yielded a slightly worse fit: $\Delta\chi^2/\Delta df = 76/30$; ΔRMSEA = .024. The next model imposed a further restriction by setting all the variances to be equal, and this model also worsened the fit: $\Delta\chi^2/\Delta df = 31/9$; ΔRMSEA = .030. The final model added a similar equality restriction in the covariances but did not alter the fit: $\Delta\chi^2/\Delta df = 4/9$; ΔRMSEA = .000.

The results from these analyses indicated that strict metric invariance does not hold across the four occasions. Given the small differences in fit and the similarity in the parameter estimates across the different measurement points, however, we considered it reasonable to assume measurement invariance. Composites of memory, space, and speed were therefore created by averaging the z scores of all the variables representing each ability at each occasion. These composites were used in all subsequent analyses.

Table 2 presents means and standard deviations for the memory, space, and speed composites across the four measurement occasions. These descriptive statistics are presented for the three age groups separately and for the overall sample. Figures 1, 2, and 3 portray the basic data used in the analyses for memory, space, and speed, respectively. The top panel in each figure portrays the complete data. In this plot, each line represents a given person's scores, and a single dot is used to represent a person with only one measurement occasion. This representation of the data combines between-persons differences (corresponding to the position on the vertical axis at which the lines begin) and within-person changes (corresponding to the trajectories of the lines), and these two components are separated in the two bottom panels. The bottom left panel was constructed by simply plotting the scores at the first measurement occasion as a function of age, and thus it represents only the between-persons age-related effects. The bottom right panel was constructed by shifting each line along the vertical axis to the zero position, and thus it represents within-person changes distinct from between-persons differences.

Figures 1, 2, and 3 show clear age differences in each cognitive ability. Across the three variables, performance seems to decrease with age, but the scores tend to increase with repeated assessments. Whereas this pattern is evident for memory, it is less apparent for space and almost nonexistent for speed, which shows individual declines across occasions. These plots illustrate the goal of our analyses, namely, to estimate the contributions of age and retest to the changes in cognitive abilities (apparent in the top panels) by considering both between-persons differences (bottom left panels) and within-person changes (bottom right panels). Details of the analytical procedures are described below.

Age-Based Mixed Growth Models

A series of models was used to investigate the changes in each cognitive ability and the effect of age and practice on such changes. A basic growth model for a dependent variable Y measured over time ($t = 1$ to T) on a person ($n = 1$ to N) can be written as

$$Y[t]_n = y_{0n} + Age[t]_n \cdot y_{sn} + e[t]_n\,, \tag{1}$$

where $Y[t]_n$ is the observed score on person n at measurement t, y_{0n} is the latent initial level score of person n, $Age[t]_n$ is the observed age of person n at measurement t, y_{sn} is a latent score of person n, representing the slope, or the individual change over time, and $e[t]_n$ is the latent error score of person n at measurement t. This model includes sources of individual

(*text continues on page 250*)

[1] The standardized factor loadings for each cognitive construct at each measurement occasion were as follows. Time 1: *Memory*—.85, .92, and .69 (for the total score across five recall trials, the delayed recall score, and the recognition test score of the RAVLT, respectively); *Spatial Ability*—.70, .61, and .59 (for the symbol digit paired associate learning test, the RCF Delayed Recall subtest, and the WAIS–R Block Design subtest, respectively); *Processing Speed*—.75, .70, .65, .78, .71, .79, and .52 (for the Digit Symbol subtest, the Trail Making Test Parts A and B, the Stroop Test A, B, and C forms, and the choice reaction time task, respectively). Time 2: *Memory*—.83, .94, and .67; *Spatial Ability*—.69, .61, and .59; *Processing Speed*—.75, .65, .58, .80, .75, .80, and .65. Time 3: *Memory*—.82, .93, and .68; *Spatial Ability*—.68, .62, and .61; *Processing Speed*—.75, .71, .63, .78, .68, .80, and .42. Time 4: *Memory*—.88, .92, and .68; *Spatial Ability*—.74, .62, and .60; *Processing Speed*—.70, .75, .71, .86, .79, .88, and .58. The correlations between the factors at the three occasions were as follows: Time 1—.65, .46, and .60, for Memory–Space, Memory–Speed, and Space–Speed, respectively; Time 2—.62, .46, and .58; Time 3—.63, .43, and .59.

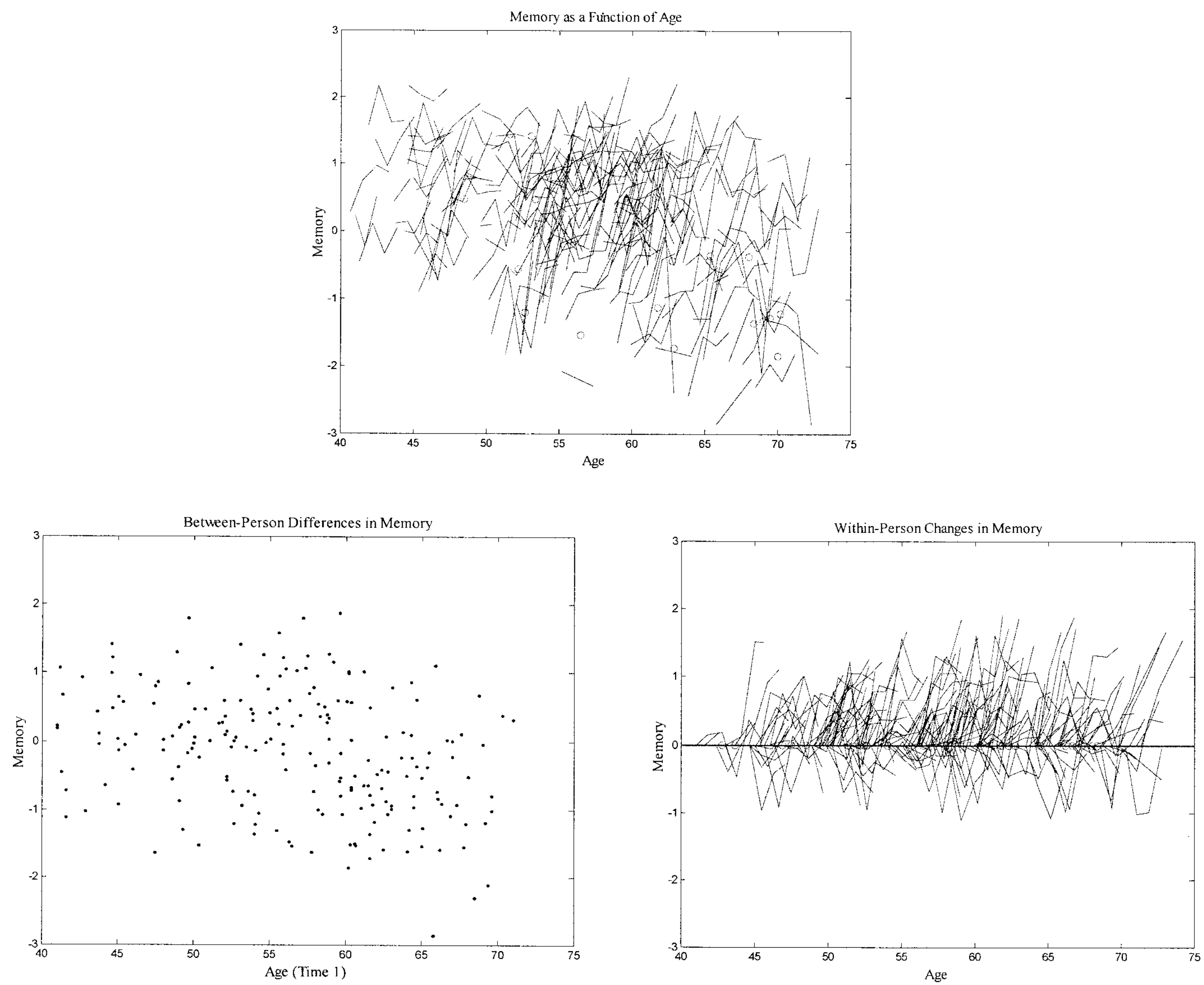

Figure 1. Memory scores as a function of age for a random 30% of all cases. Top: Individual score trajectories over age. Each line represents an individual's multiple scores; circles represent those individuals with only one data point. Bottom left: Scores at first measurement occasion. Bottom right: Within-person changes. The horizontal line is included to distinguish positive change (above the line) from negative change (below the line).

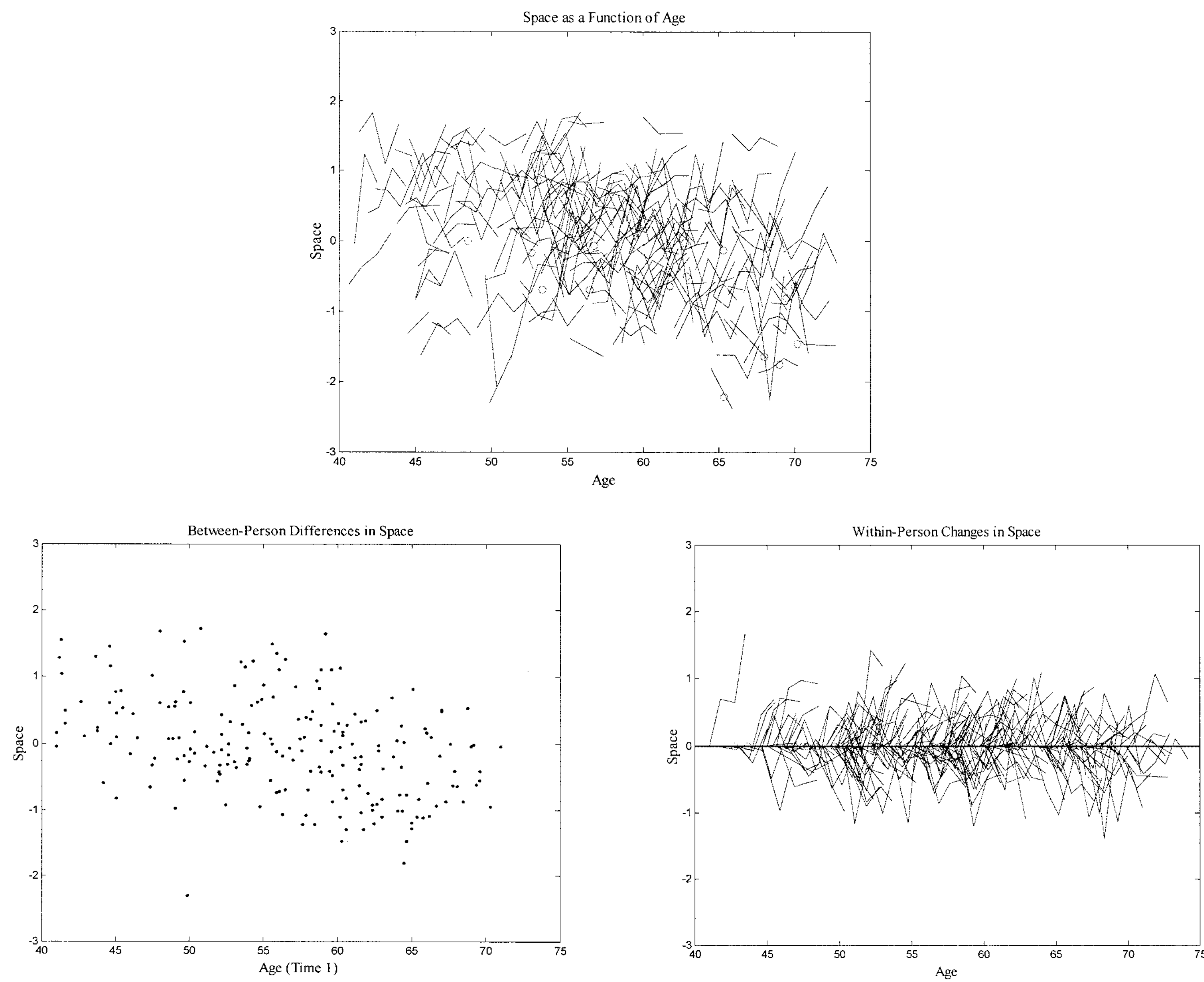

Figure 2. Space scores as a function of age for a random 30% of all cases. Top: Individual score trajectories over age. Each line represents an individual's multiple scores; circles represent those individuals with only one data point. Bottom left: Scores at first measurement occasion. Bottom right: Within-person changes. The horizontal line is included to distinguish positive change (above the line) from negative change (below the line).

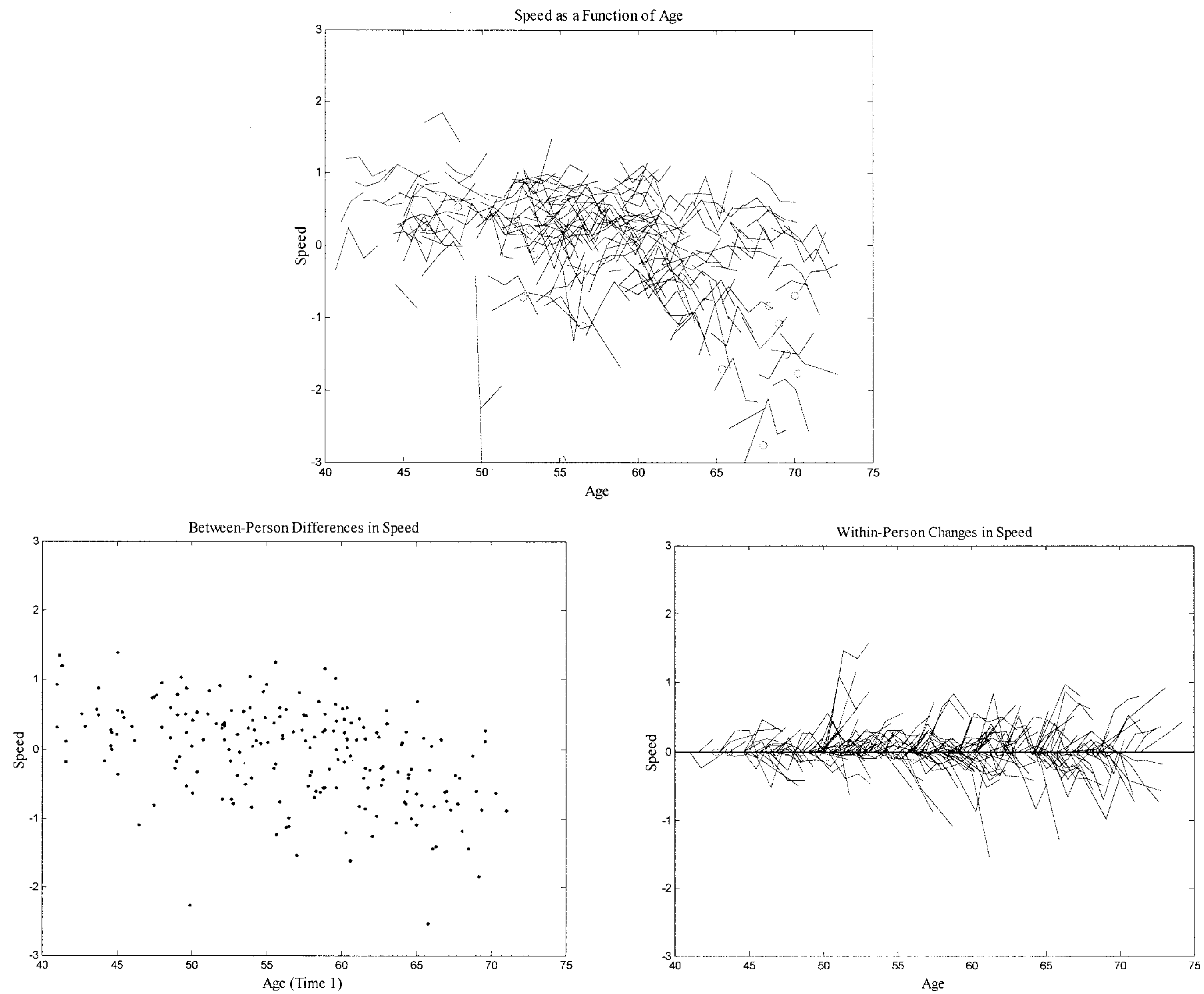

Figure 3. Speed scores as a function of age for a random 30% of all cases. Top: Individual score trajectories over age. Each line represents an individual's multiple scores; circles represent those individuals with only one data point. Bottom left: Scores at first measurement occasion. Bottom right: Within-person changes. The horizontal line is included to distinguish positive change (above the line) from negative change (below the line).

differences in the level and slope, whose terms can be decomposed at a second level as

$$y_{0n} = \mu_0 + e_{0n},$$
$$y_{sn} = \mu_s + e_{sn}, \tag{2}$$

where the level and slope scores have fixed group means (μ_0 and μ_s) and residuals (e_{0n} and e_{sn}), and these residuals have variance components (σ_0^2, σ_s^2, and σ_{0s}). In the case of cross-sectional data, this same model can be written as

$$Y[1]_n = y_{0n} + Age[1]_n \cdot y_s + e[1]_n, \tag{3}$$

where $Y[t]_n$ is the observed score on person n at the only measurement occasion (1), y_{0n} is still the latent initial level score of person n, and $Age[1]_n \cdot y_s$ becomes the difference in the dependent variable Y as a function of 1 unit of cross-sectional age (i.e., 1 year). The basic model of Equation 1 can take different forms depending on the researcher's hypotheses. For example, in the case of a hypothesis of no growth ($y_s = 0$), this basic model can be rewritten as

$$Y[t]_n = y_{0n} + e[t]_n. \tag{4}$$

An alternative model is a polynomial model that considers different nonlinear age functions (i.e., quadratic, cubic), as

$$Y[t]_n = y_{0n} + Age[t]_n \cdot y_{s1n} + Age[t]_n^2 \cdot y_{s2n} + e[t]_n, \tag{5}$$

where $Age[t]_n^p$ is the age basis of power p for person n, and y_{spn} is the latent polynomial component score of person n.

Age- and Occasion-Based Mixed Growth Models

The initial age-based model expressed in Equation 1 can also be extended to incorporate exogenous variables. In the current investigation, we are interested in examining the effects of practice, and thus the model can be written as

$$Y[t]_n = y_{0n} + Age[t]_n \cdot y_{sn} + P[t] \cdot p_{sn} + e[t]_n, \tag{6}$$

where $P[t]$ represents the practice, or retest, effects of person n at measurement t, and p_{sn} represents the slope, or individual change in practice over time, for person n. The other terms are as specified in Equation 1. Similarly to previous specifications, the terms of this model can be decomposed at a second level as

$$y_{0n} = \mu_0 + e_{0n},$$
$$y_{sn} = \mu_s + e_{sn},$$
$$p_{sn} = \mu_p + e_{pn}, \tag{7}$$

where the level, slope, and practice scores have fixed group means (μ_0, μ_s, and μ_p, respectively) and residuals (e_{0n}, e_{sn}, and e_{pn}), and these residuals have variance components (σ_0^2, σ_s^2, and σ_p^2) and can co-vary among themselves (σ_{0s}, σ_{0p}, and σ_{sp}). Depending on data conditions, restrictions could be imposed in some of these components (e.g., if there are only two repeated observations per person, the covariances between practice and other terms could be set to zero). According to this model, change in Y can be described as a function of two processes that unfold over time: age and practice. $Age[t]_n \cdot y_{sn}$ can vary over time for each person, so this term represents an age-based growth process. That is, it is a slope based on age at testing occasion t (i.e., average change in Y per year for a person n) and takes into consideration the time dependency of scores for each person. In turn, the practice term, $P[t] \cdot p_{sn}$, reflects a growth process based on the measurement occasion (i.e., average change in Y per unit change in retest for a person n). A path diagram of this model is depicted in Figure 4. This figure represents a variable Y, measured at t occasions, whose growth is modeled as a function of an intercept y_0 and two slopes y_s and p, representing processes related to age and retest. An important feature of this figure—and of the underlying model—is the basis coefficients (i.e., paths from age and retest to the observed variables). By setting these paths to fixed or latent values, one can test different hypotheses of change.

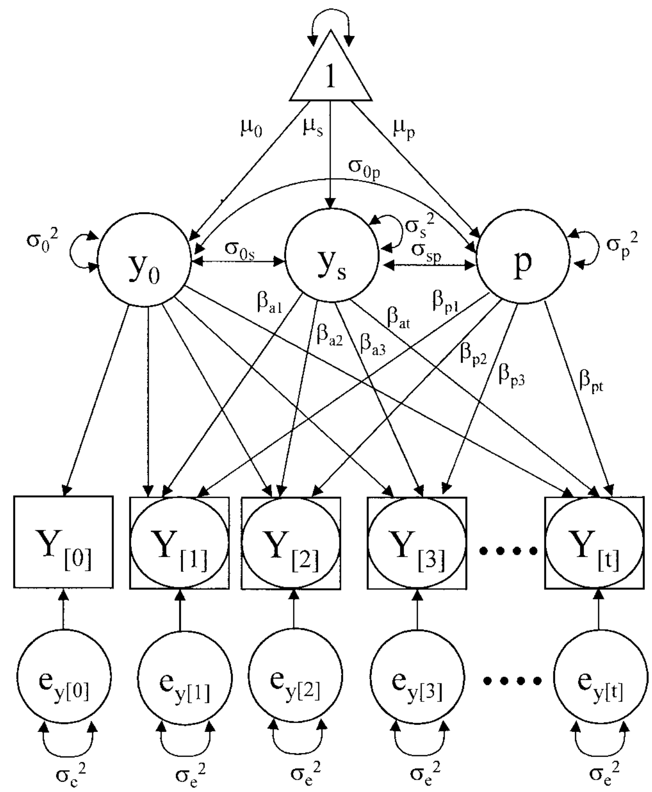

Figure 4. Path diagram of a latent growth model with two processes. $Y_{[t]}$ = score at time t; y_0 = intercept; y_s = "age" slope; p = "retest" slope; $e_{y[t]}$ = uniquenesses; 1 = constant; β_a = basis coefficients for age; β_p = basis coefficients for practice; μ = means; σ^2 = variances; σ = covariances.

Separating the effects of age and practice requires a model with a mixture of both age and practice processes. Examining the characteristics of either process can be accomplished by modeling specific growth hypotheses using basis coefficients (McArdle & Anderson, 1990; McArdle & Woodcock, 1997). For example, one could fit a model positing linear processes for both age and practice by setting the age basis coefficients as $A[t]_n$ fixed = 0, 1, 2, . . . age (t), and the practice basis coefficients as $P[t]$ fixed = 0, 1, 2, . . . occasion (t). This model would represent a function that changes linearly with age and is affected by a practice influence, which takes place at the second assessment and increases linearly after that. Alternative basis coefficients could be used to test different hypotheses. For example, considering the four measurement occasions in the current data, the specification $P[t]$ = 0, 1, 1, 1 would test for practice effects that take place at the first retest and remain constant thereafter. In contrast, the specification $P[t]$ = 0, 1, 4, 9 would test for a quadratic practice effect. To examine more complex growth functions, some of these coefficients can be freed so as to represent unequal, or latent, practice growth. For example, the specification $P[t]$ = 0, ?, ?, 1 would allow examination of whether practice effects are discontinuous across occasions.

The model expressed in Equation 6 could also include a Practice × Age interaction. Such an interaction term would be helpful for examining the relation of practice to age (i.e., whether individuals benefit differently from retest as a function of their age). A model with an interaction term could be written as

$$Y[t]_n = y_{0n} + Age[t]_n \cdot y_{sn} + P[t] \cdot p_{sn} + Age \cdot P[t] \cdot y_{apn} + e[t]_n , \tag{8}$$

where the interaction term $Age \cdot P$ represents the extent to which practice effects differ as a function of age. To avoid overparameterization of this model, the random effects associated with the interaction term (i.e., variance of the interaction and covariance with all the other terms in the equation) can be excluded.

Age- and Occasion-Based Mixed Growth Models Across Groups

One approach to testing for differences in age and retest effects across age groups is to fit the same analytic model to data from different age groups. That is, the model in Equation 6 could be written as

$$Y[t]_n^{(g)} = y_{0n}^{(g)} + Age[t]_n^{(g)} \cdot y_{sn}^{(g)} + P[t]^{(g)} \cdot p_{sn}^{(g)} + e[t]_n^{(g)} , \tag{9}$$

where $g = 1$ to G represents the group, and thus differences in the growth processes across individuals of different ages could be examined. If evidence of a Practice × Age interaction was found in previous analyses, it could be decomposed using this approach as well. In addition, the mixed models presented here rely on a convergence assumption (after Bell, 1953). This assumption can be formally tested using a multiple-group analysis in which one examines whether the growth of a selected cognitive variable can be described using the same functions (i.e., age and retest) for all ages.

Results

Age-Based Mixed Growth Analyses

The analyses followed the logic outlined in the Method section and were conducted using the procedures MIXED and NLMIXED in SAS (Littell, Miliken, Stoup, & Wolfinger, 1996). These procedures use a maximum likelihood algorithm to generate parameter estimates using all available data. In the case of incomplete data, this approach assumes they are missing at random, an assumption that may not be justified. As noted earlier, however, the analyses were repeated with the data from participants who completed testing at all four occasions, and the results indicated similar trends.

As a starting point, we first examined a cross-sectional model based on the data at the first occasion. This model, expressed in Equation 3, yielded an intercept, a cross-sectional slope for age (centered at 39.9 years, the youngest age), and a residual. The parameter estimates from this model are displayed in the first columns of Tables 3, 4, and 5 (for memory, space, and speed, respectively) and simply describe the relation of age and cognitive abilities if data were available at one occasion only. These parameters represent the between-persons differences portrayed in the bottom left panels of Figures 1, 2, and 3. Of particular interest here is the cross-sectional slope, which reflects the mean difference in the outcome across individuals who differ in age by 1 year. The estimates were similar across the three variables and indicated a small and similar age-related decline in cognition ($\mu_a = -.035$,

Table 3
Age-Based Mixed Growth Model Parameters for Memory

Parameter	Age $[t_0]$ cross-sectional	Age $[t_t]$ longitudinal	Age $[t_t]$ + 1st retest	Age $[t_t]$ + linear retest	Age $[t_t]$ + latent practice
		Fixed effects			
μ_0 Intercept	.581 (8)	.281 (4)	.381 (6)	.604 (9)	.573 (9)
μ_a Linear age	−.035 (9)	−.002 (0)	−.023 (7)	−.035 (10)	−.035 (10)
μ_p Practice			.409 (21)	.619 (24)	.619 (23)
$\beta_{p[1]}$			0 (=)	0 (=)	0 (=)
$\beta_{p[2]}$			1 (=)	.33 (=)	.482 (18)
$\beta_{p[3]}$			1 (=)	.66 (=)	.731 (24)
$\beta_{p[4]}$			1 (=)	1 (=)	1 (=)
		Random effects			
σ_0^2 Intercept	.000 (?)	.731 (5)	.700 (5)	.647 (5)	.659 (5)
σ_a^2 Linear age		.001 (3)	.001 (1)	.001 (1)	.001 (1)
σ_p^2 Practice			.032 (1)	.090 (3)	.114 (4)
σ_{0a} Level–Age		−.018 (2)	−.011 (1)	−.008 (1)	−.008 (1)
σ_{0p} Level–Practice			−.097 (3)	−.098 (2)	−.113 (2)
σ_{ap} Age–Practice			.002 (1)	.002 (1)	.002 (1)
σ_e^2 Residual variance	.704 (20)	.200 (28)	.156 (23)	.137 (24)	.131 (24)
		Goodness of fit			
Likelihood (−2 log-likelihood)	2,061	5,162	4,716	4,542	4,511
No. of parameters	4	6	10	10	12

Note. Values in parentheses indicate t (for fixed effects) or z (for random effects) values in absolute terms. (=) indicates a fixed parameter. (?) indicates a parameter whose standard errors were not identified. All parameters are maximum-likelihood estimates from SAS PROC MIXED and NLMIXED.

Table 4
Age-Based Mixed Growth Model Parameters for Space

Parameter	Age [t_0] cross-sectional	Age [t_t] longitudinal	Age [t_t] + 1st retest	Age [t_t] + linear retest	Age [t_t] + latent practice
		Fixed effects			
μ_0 Intercept	.596 (10)	.450 (8)	.508 (9)	.644 (11)	.670 (11)
μ_a Linear age	−.036 (11)	−.023 (8)	−.031 (10)	−.040 (13)	−.040 (13)
μ_p Practice			.118 (7)	.258 (13)	.239 (11)
$\beta_{p[1]}$			0 (=)	0 (=)	0 (=)
$\beta_{p[2]}$			1 (=)	.33 (=)	.111 (1)
$\beta_{p[3]}$			1 (=)	.66 (=)	.869 (11)
$\beta_{p[4]}$			1 (=)	1 (=)	1 (=)
		Random effects			
σ_0^2 Intercept	.000 (?)	.374 (7)	.387 (6)	.451 (4)	.517 (4)
σ_a^2 Linear age		.000 (0)	.000 (0)	.000 (0)	.000 (0)
σ_p^2 Practice			.007 (0)	.012 (1)	.026 (1)
σ_{0a} Level–Age		.002 (1)	.001 (0)	−.002 (0)	−.006 (1)
σ_{0p} Level–Practice			.014 (1)	.024 (1)	−.022 (1)
σ_{ap} Age–Practice			.007 (1)	.000 (0)	.001 (1)
σ_e^2 Residual variance	.516 (20)	.117 (30)	.114 (23)	.108 (25)	.102 (25)
		Goodness of fit			
Likelihood (−2 log-likelihood)	1,813	3,940	3,881	3,772	3,746
No. of parameters	4	6	10	10	12

Note. Values in parentheses indicate t (for fixed effects) or z (for random effects) values in absolute terms. (=) indicates a fixed parameter. (?) indicates a parameter whose standard errors were not identified. All parameters are maximum-likelihood estimates from SAS PROC MIXED and NLMIXED.

−.036, and −.043, for memory, space, and speed, respectively). In addition, the residual term suggests there is a substantial amount of residual variance unexplained by the model (σ_e^2 = .704, .516, and .463, respectively), which is largest for memory and smallest for speed.

In the next set of analyses, we examined a series of age-based mixed models (i.e., including both within-person and between-persons information) to identify the age function that best described the different cognitive abilities over time (i.e., modeling data in the top panels of Figures 1, 2, and 3). These models included a model of no growth, a linear age model (with age centered at 39.9, the youngest age), and a quadratic age model. Compared with a model of no growth, adding a linear age term (Equation 1) improved the fit substantially for both space and speed ($\Delta\chi^2/\Delta df$ = 206/3 and 111/3, respectively). This, however, was not true for memory ($\Delta\chi^2/\Delta df$ = 3/3). In contrast to the cross-sectional model, the estimates for memory from this linear model suggest a flat trajectory over age ($\mu_a \approx 0$) with almost no variation across persons (σ_a^2 = .001) and a large residual (σ_e^2 = .200) representing unexplained within-individual variation. The lack of age-related effects in this case reflects the opposite patterns of within-person change (positive, in bottom right panel of Figure 1) and between-persons differences (negative, in bottom left panel of Figure 1), which cancel one another out.

For the space and speed variables, the estimates from the linear age model are similar to the cross-sectional estimates, reflecting weaker within-person changes (as in the bottom left panels of Figures 2 and 3). Although the age estimates for both variables are reduced in magnitude (μ_a = −.023 and −.033 for space and speed, respectively), they still indicate declines in cognition over age with small (for speed) or no detectable (for space) variation across people. These results are presented in the second columns of Tables 3, 4, and 5.

For the three variables, a quadratic relation of age was detectable, although with a small parameter estimate (μ_a^2 = −.001; estimates not presented in tables). When compared with a linear model, this quadratic model yielded a better fit for the memory and speed variables, ($\Delta\chi^2/\Delta df$ = 152/4 and 210/4, respectively) but not for the space variable ($\Delta\chi^2/\Delta df$ = 7/4). These results indicate that there is a small decline in cognitive abilities over age, which seems to accelerate as individuals get older. For the space variable, however, the decline seems to be uniform over age. Although informative, these results do not take into account retest effects, so the potential within-person changes that are due to such effects are incorporated into the age effects. That is, ignoring practice effects may be biasing the estimates of the age effects. In the next set of analyses, we incorporated both terms in the models to examine their separate contributions.

Age- and Occasion-Based Mixed Growth Analyses

The results from the analyses estimating both age and retest effects are presented in the last three columns of Tables 3, 4, and 5. The first model posited a linear function for age and practice effects that take place at the first retest and remain constant thereafter (basis $P[t]$ = 0, 1, 1, 1). When comparing the fit of this model with a linear model without practice (i.e., Δ−2 log-

Table 5
Age-Based Mixed Growth Model Parameters for Speed

Parameter	Age [t_0] cross-sectional	Age [t_t] longitudinal	Age [t_t] + 1st retest	Age [t_t] + linear retest	Age [t_t] + latent practice
Fixed effects					
μ_0 Intercept	.696 (12)	.604 (13)	.671 (14)	.661 (13)	.681 (13)
μ_a Linear age	−.043 (14)	−.033 (12)	−.040 (14)	−.037 (11)	−.040 (13)
μ_p Practice			.089 (6)	.062 (3)	.087 (4)
$\beta_{p[1]}$			0 (=)	0 (=)	0 (=)
$\beta_{p[2]}$			1 (=)	.33 (=)	.477 (24)
$\beta_{p[3]}$			1 (=)	.66 (=)	.700 (19)
$\beta_{p[4]}$			1 (=)	1 (=)	1 (=)
Random effects					
σ_0^2 Intercept	.000 (?)	.262 (4)	.302 (5)	.421 (4)	.347 (4)
σ_a^2 Linear age		.001 (4)	.001 (3)	.002 (3)	.001 (3)
σ_p^2 Practice			.067 (6)	.087 (7)	.247 (10)
σ_{0a} Level–Age		−.005 (1)	−.007 (1)	−.015 (2)	−.010 (1)
σ_{0p} Level–Practice			−.016 (1)	−.108 (3)	.112 (3)
σ_{ap} Age–Practice			.001 (1)	.008 (4)	−.008 (3)
σ_e^2 Residual variance	.463 (20)	.082 (29)	.063 (23)	.053 (24)	.050 (23)
Goodness of fit					
Likelihood (−2 log-likelihood)	1,723	3,280	3,252	3,125	2,999
No. of parameters	4	6	10	10	12

Note. Values in parentheses indicate *t* (for fixed effects) or *z* (for random effects) values in absolute terms. (=) indicates a fixed parameter. (?) indicates a parameter whose standard errors were not identified. All parameters are maximum-likelihood estimates from SAS PROC MIXED and NLMIXED.

likelihood) there was a substantial improvement in fit ($\Delta\chi^2/\Delta df$ = 446/4, 59/4, and 28/4, for memory, space, and speed, respectively). The estimates from this model indicate age-related declines for all variables ($\mu_a = -.023$, $-.031$, and $-.040$, respectively) with very small variation across individuals. With each variable, there is a detectable improvement in performance at the first retest (μ_p = .409, .118, and .089, respectively), with individual variation for speed only ($\sigma_p^2 = .067$). Compared with a model with no retest, this model reduced the residual variance for all variables, although this was more apparent for memory, the variable with the largest within-person changes.

The next model tested linear functions for both age and practice. That is, it extended the previous model by examining whether retest effects occurred after the second measurement occasion. This model resulted in improved fit for all variables ($\Delta\chi^2/\Delta df$ = 174/0, 109/0, and 127/0, respectively), suggesting that retest effects took place not only at the second occasion but also at subsequent assessments. To examine whether the fit improvement was primarily due to effects of practice on the mean or to effects on the between-persons variability, we entered these terms separately. For memory and space, the improvement was due primarily to the mean ($\Delta\chi^2/\Delta df$ = 593/1, 155/1), with a minor contribution by the variance terms ($\Delta\chi^2/\Delta df$ = 21/3, 7/3). For speed, however, the improvement was mostly due to the variance components ($\Delta\chi^2/\Delta df$ = 240/3) and not so much due to the mean ($\Delta\chi^2/\Delta df$ = 9/1), suggesting that although the overall retest mean was small, there was substantial individual variation around it. Finally, we examined whether adding a quadratic age function to the model improved the fit further. This was the case for speed ($\Delta\chi^2/\Delta df$ = 25/5) but not for memory or space ($\Delta\chi^2/\Delta df$ = 3/5 and 2/5, respectively). Moreover, for all three variables, the estimate for a quadratic age effect was near zero ($\mu_a^2 = -.001$). On the basis of these results and a desire to have comparable analyses across the three cognitive variables, we decided not to examine further models with quadratic age effects.

The results from a model with linear terms for age and retest are displayed in the fourth columns of Tables 3, 4, and 5. For memory (see Table 3), these results yielded an average initial score (μ_a = .604) at 39.9 years (the point at which age is centered), with variation around this estimate ($\sigma_0^2 = .647$). That is, the initial score for 95% of individuals ranged from −1.01 to 2.21 (i.e., .604 ± 2 $\sqrt{.647}$). The estimate for age ($\mu_a = -.035$) indicates that, on average, individuals' performance in memory declines by this amount per year, without apparent variation across individuals ($\sigma_a^2 \approx 0$). The estimate for practice ($\mu_p = .619$) suggests that as individuals repeat the memory assessment, their performance increases. The coding for the linear practice effect was $P[t]$ = 0, .33, .66, 1, so this estimate represents an overall improvement that is uniform across the four occasions. Although not large, some individual variation in this effect is detectable ($\sigma_p^2 = .090$). The only significant covariance is between initial level and practice ($\sigma_{0p} = -.098$), suggesting that the retest effect is smaller for those individuals who started at higher initial levels.

The next model considered a latent practice term, representing practice effects that are unequal across occasions. For this purpose, we relaxed the practice loadings at the second and third occasions ($P[t] = 0, ?, ?, 1$). Relaxing these two parameters improved the fit ($\Delta\chi^2/\Delta df = 31/2$; ΔRMSEA = .133), which suggests that the practice effects for memory are not equal across measurement occasions. The parameter estimates from this model confirmed this point. Although the overall practice effect across all occasions was the same ($\mu_p = .619$) as in previous models, the practice loadings indicated that about 48% of this effect takes place at the second assessment ($\beta_{p[2]} = .482$), with smaller effects at the third and fourth occasions (25% and 27%, respectively). By comparing the results from this model with those from a baseline model (i.e., a random intercept model), one can obtain the proportions of between- and within-person variance that are explained by this model (see Raudenbush & Bryk, 2002). For memory, such proportions are 18% and 32%, respectively.

The same sets of analyses were conducted for the space and speed variables, and the respective results are presented in Tables 4 and 5. For both cognitive abilities, these analyses yielded negative age effects and positive practice effects. Whereas the former effects were similar in magnitude to the effects found for the memory variable, the latter effects were smaller for space and close to zero for speed, reflecting near symmetry of lines above and below zero in the bottom right panels of Figures 2 and 3. Also for both abilities, a model with a nonlinear practice effect yielded a better fit than a linear practice specification, and this improvement was much more apparent for speed ($\Delta\chi^2/\Delta df = 26/2$, ΔRMSEA = .121 for space; $\Delta\chi^2/\Delta df = 126/2$, ΔRMSEA = .275 for speed). For space, the estimates for a linear age and latent practice model indicate that performance decreases slightly as a function of age ($\mu_a = -.040$), without perceptible variation across persons ($\sigma_a^2 \approx 0$). Moreover, there is an overall practice effect across the four occasions ($\mu_p = .239$), which does not seem to vary across individuals, and most of this effect takes place at the third occasion ($\beta_{p[3]} = .869$; 76%). The estimated percentages of between- and within-person variance explained by this model are 5% and 18%, respectively.

The interpretation of the estimates for speed is similar (see Table 5). Performance in this cognitive ability shows a decrease over age ($\mu_a = -.040$), with a very small variation across persons ($\sigma_a^2 = .001$). The overall practice effect is small ($\mu_p = .087$), but there is substantial variation across persons in this effect ($\sigma_p^2 = .247$). Furthermore, about 48% of the practice-related improvement seems to occur at the first retest ($\beta_{p[2]} = .477$), whereas about 22% and 30% take place at the third and fourth occasions, respectively. Compared with a baseline model, this model explains 26% and 39% of the between- and within-person variance, respectively.

To investigate whether individuals benefit differently from practice depending on their age, we examined a possible Age $\times$ Practice interaction as represented in Equation 8. A model including an Age $\times$ Practice interaction did not improve the fit of a model without the interaction for any of the abilities. Moreover, the estimate for the interaction term was not different from zero for memory ($\mu_{int} \approx 0$) and was very small for space and speed ($\mu_{int} = -.007$ and $-.010$, respectively). This result suggests that there are similar retest effects across the age range from 40 to 74 years for these variables. This finding can be seen in the bottom right panels of Figures 1, 2, and 3, in which the within-person changes appear to be similar across the age range.

Age- and Occasion-Based Mixed Growth Models Across Groups

To identify possible group differences, we divided the sample into three age groups ($<$ 50 years, 50–59 years, and $>$ 60 years) and fitted a series of models to examine invariance across them. The first model constrained all the parameters to be equal across the groups (i.e., fixed to the values for the pooled sample). This model tested the extent to which the linear age and latent practice components found for the overall sample held across the three groups. Subsequent models relaxed parameters to be free across groups until all five fixed effects (i.e., intercept, age slope, practice slope, and two basis coefficients) were allowed to differ across groups. For memory, this least restrictive model did not improve on the fit of a full invariant model ($\Delta\chi^2/\Delta df = 21/15$), with the relative misfit per group ($\Delta\chi^2/\Delta df = 3/5$, 13/5, and 5/5, respectively) indicating a minor improvement for the second group only. Similar invariant results were found for space ($\Delta\chi^2/\Delta df = 14/15$), with no differences in misfit across groups. For speed, however, invariance across age groups did not hold ($\Delta\chi^2/\Delta df = 118/15$), with improvement in fit being detectable for all age groups ($\Delta\chi^2/\Delta df = 60/5$, 21/5, and 37/5, respectively). Additional analyses using less restrictive models indicated differences in the practice variance across groups but not in the variance terms for the intercept or the age effects. Estimates from this least restrictive model are presented in Table 6.

The results from the multiple-group analyses suggest that for the memory and space variables, the results obtained from the overall sample apply to all age groups. For speed, however, this does not seem to be true, and differences in fixed and random effects are detectable across groups. Such differences indicate that age effects are most pronounced for older individuals, reflecting a quadratic age function, and retest effects are most apparent for younger

Table 6
Estimates From a Linear-Age and Latent-Practice Model Across Age Groups for Speed

Parameter	$<$ 50 years	50–59 years	$>$ 60 years
μ_{00} Intercept	.452 (6)	.360 (2)	1.110 (3)
μ_{10} Linear age	$-.019$ (1)	$-.012$ (1)	$-.061$ (4)
μ_{01} Latent practice	.082 (2)	.010 (0)	.077 (1)
$\beta_{p[1]}$	0 (=)	0 (=)	0 (=)
$\beta_{p[2]}$	1.164 (8)	.481 (16)	.232 (4)
$\beta_{p[3]}$	.877 (7)	.591 (11)	.729 (12)
$\beta_{p[4]}$	1 (=)	1 (=)	1 (=)
σ_p^2 Practice	.128 (4)	.291 (11)	.256 (8)
σ_e^2 Residual variance	.039 (9)	.035 (16)	.065 (15)
-2 log-likelihood	516	1026	1268
$\Delta\chi^2/\Delta df$[a]	69/6	24/6	37/6

Note. Values in parentheses indicate t values in absolute terms. (=) indicates a fixed parameter. Random parameters are not shown but were fixed at values for the overall sample (as in Table 5). Sample sizes are as follows: Group A = 181, Group B = 357, Group C = 289. $\beta_{p[1]} = 0$ and $\beta_{p[4]} = 1$ are fixed values. All parameters are maximum likelihood estimates from SAS PROC NLMIXED.
[a] $\Delta\chi^2/\Delta df$ in relation to a full invariant model.

individuals, with most effects occurring at the first retest (i.e., $\beta_{p[2]} = 1.16$, indicating that the retest effect at this occasion exceeded the value at the final occasion, $\beta_{p[4]} = 1$). Although no average retest effects are apparent for the older groups, such effects exist for some older individuals, and these effects seem to materialize later than those for the younger individuals. The results from this last set of analyses can be translated into curves for each variable. Such latent curves—with 95% confidence intervals—are plotted in Figure 5 and represent the expected trajectory for each variable and age group based on the combined effects of age and retest.

Discussion

Summary of Findings

In the current report we examined the possibility of modeling age and retest as separate processes underlying changes in cognitive abilities. The findings of our analyses indicate that both age and retest should be modeled simultaneously when analyzing longitudinal data in which practice effects may occur because of repeated assessments. Under these conditions, ignoring the retest process in the model tends to underestimate the age effects, with more severe bias occurring when retest effects are larger. In such models, the age estimates need to capture all changes that are due to both age and retest, and if the retest effects are positive, the true age effects will be underestimated.

We illustrated these effects in our analyses by fitting various models to three different composite variables. In the case of the memory variable, there was substantial improvement with retest, and thus the age estimate was negligible when the analytic model did not include retest effects. When such effects were part of the model, however, the age effects were negative and were accompanied by positive retest effects. In other words, this model indicated that memory declines with increased age but that these effects can be masked by improvement that takes place with repeated assessments. The difference in the age estimates between models with and without retest was much smaller for the space variable and almost trivial for the speed variable, because the retest effects were much smaller. Rabbitt et al. (2001) cautioned about longitudinal studies in which practice effects are ignored because, they argued, the true rates of cognitive decline are underestimated. Our findings are in line with this contention and further show that the bias will depend on the magnitude of the omitted practice effects, which can vary across variables.

There are several possible causes of the retest effects found in these analyses. First, given that the battery was administered in exactly the same way at all occasions, it is likely that part of the retest effects was due to the fact that some of the test items were actually remembered after the first test occasion. Second, more general factors such as increased familiarity with the testing situation and/or decreased test anxiety could also be responsible for the retest effects. Such general factors would presumably affect retest scores on a memory test even if the test was repeated with a new set of items, and general factors may be the main source of retest effects on speed tests in which familiarity with the test items should have less of an effect. The discovery of greater retest effects for memory may be due to a combination of item-specific and general effects.

One goal of our analyses, in addition to separating age and retest processes, was to examine the relative impact of retest at each measurement occasion. For all the variables, our results indicate that practice took place not only at the first retest but at subsequent assessments as well, with unequal improvements across intervals. For the memory and speed variables, about half of all the retest effects occurred at the second assessment, with the rest of the effects distributed across the third and fourth assessments. For the space variable, however, retest effects were most apparent at the third measurement occasion (i.e., about 87%), with some additional effects at the fourth assessment. These results showing practice effects across the four measurement occasions are in line with the Rabbitt et al. (2001) findings in which performance in a fluid ability measure improved over the course of four assessments, independently from negative age effects. Like the case for the memory and speed variables in our analyses, in the Rabbitt et al. study, about half of the practice effects occurred at the second assessment, with perceptible but decreasing effects across the third and fourth occasions.

Our findings are also similar to results reported by Wilson et al. (2002) in a study of cognitive decline among older individuals. In the Wilson et al. study, individuals' performance in several cognitive domains improved at the first retest, with larger improvements for word generation. Improvement at the second retest was also found for measures of word generation, perceptual speed, and visuospatial ability, and an additional increment at the third retest was found for word generation and perceptual speed. Also similar to our findings, when a term for practice effects was included in the Wilson et al. analyses, the estimates of negative age effects increased for all measures. Our results, however, depart somewhat from the results of other important longitudinal studies reporting smaller practice effects (Hultsch, Hertzog, Dixon, & Small, 1998; Schaie, 1988). Possible reasons for these differences are the selected and more homogeneous sample in the current study (i.e., men between 40 and 70 years of age) and the smaller retest intervals.

To test for differences in age and retest effects across age groups, we conducted separate analyses for different groups. These analyses were helpful in identifying how age and practice processes that underlie cognitive changes differ as a function of age. The findings here revealed that for memory and space, both age and retest effects were similar across different age groups. For speed, however, pronounced age effects were found for individuals 60 years old and older, and practice effects were apparent for younger individuals (i.e., individual variation in retest was present at all ages). It is unclear whether the current results would hold if the retest intervals were expanded or the age range of the participants was extended. For example, it is possible that with longer retest intervals, retest effects that reflect item-specific influences would decrease, especially for older adults. Another possibility is that with longer retest intervals, retest effects that reflect learning through repetition would decrease, especially for older adults. In addition, the relatively low mean age of the adults in the oldest group (67.5 years at Time 4) could also have biased the findings toward comparable retest effects across age groups. In line with the Rabbitt et al. findings (2001), our results suggest that benefits from retest that are likely due to learning through repetition apply to persons of all ages, even to those who show a marked cognitive decline. If possible retest effects are likely to differ by age, we

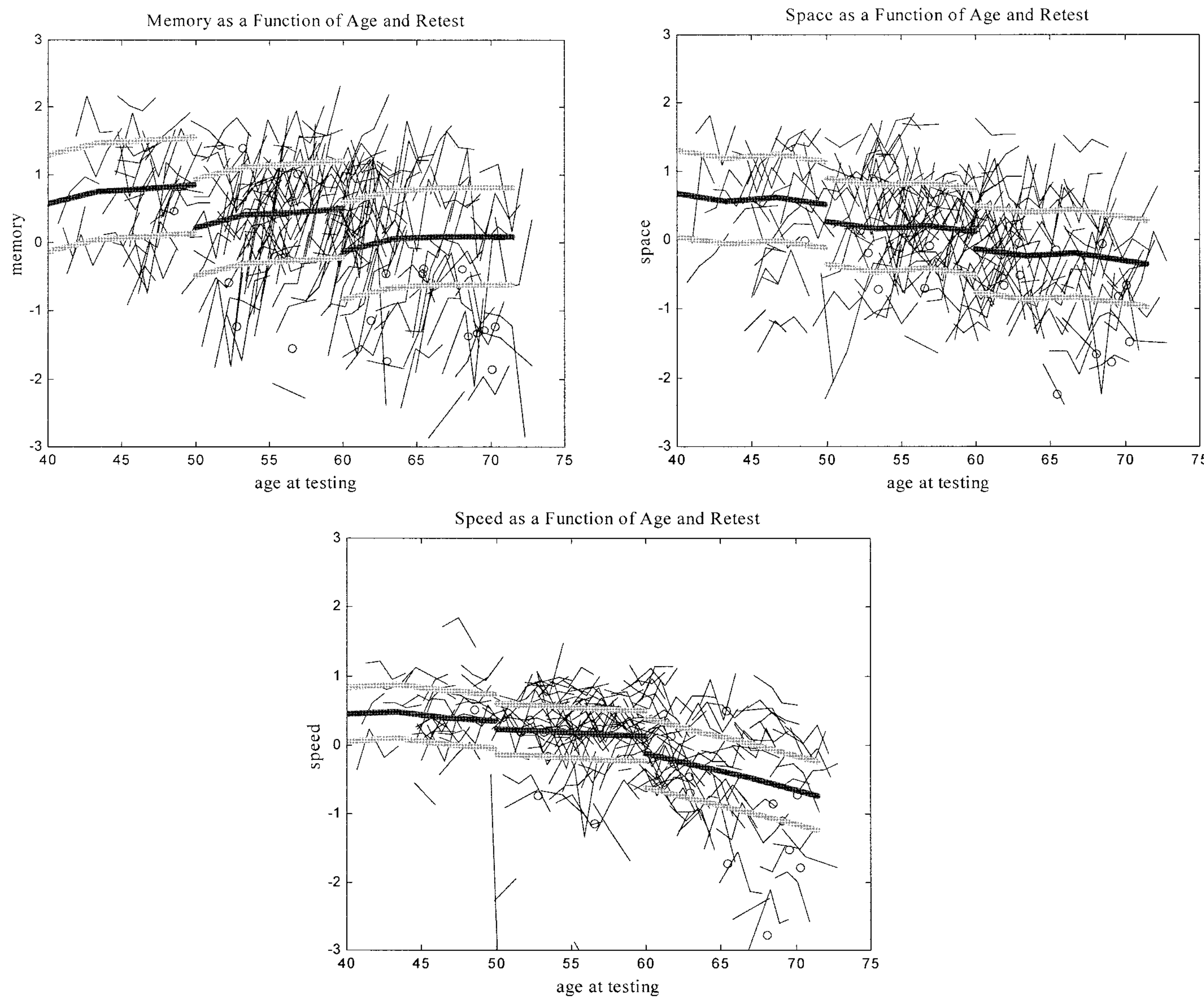

Figure 5. Expected trajectories (darker, middle line) and 95% confidence intervals (lighter, gray external lines) for each cognitive ability as a function of age and retest for the three age groups. The background data represent a random 30% of all cases.

suggest that an Age × Retest interaction be included in the analytic model. If it is found that younger adults benefit from retests more than do older adults, then correcting for retest effects may reduce the magnitude of the age differences because the extra advantage of younger adults is eliminated.

The results from our models indicate that there was a small variation in the slope for age across persons and variables. That is, based on our analyses, all individuals seemed to follow a similar age-related rate of decline. This was true even for models that did not include a term for retest effects, which suggests that the variation in retest did not absorb the variance in age. It is likely that this lack of variation was caused by the relatively short retest intervals and the relatively homogeneous sample. This interpretation would also help explain why the retest estimates were larger in studies that involved individuals of similar ages but larger retest intervals (e.g., McArdle & Anderson, 1990; Rabbitt et al., 2001).

Methodological Issues

Our main purpose in this report was to model separate age and retest effects in studies involving repeated assessments. This separation is difficult in the presence of a high correlation between the increment in age and the increment in retest occasions. Possible remedies for this confounding are a large variation in retest and a wide age range. The retest intervals in this study were close to 1 year, so the data are not optimal for the unequivocal separation of age and retest effects. The age range, however, was wide, ranging from 40 to 74 years. This feature allowed us to weaken the age–retest correlation and to model separate processes for age and retest. In the raw data, the correlation between age and retest occasion across persons and variables was .80. In the analyses, however, the correlations between estimates of age effects and effects associated with successive retests were significant in some models of processing speed only.

In addition to separating age and retest effects, we were able to model individual variation in both age and retest without imposing constraints in the models, although the estimates of individual differences in age-related effects were small. The addition of practice to the model, including its random component, did not alter the variance term for age, so all age effects represented longitudinal slopes with variation across individuals. It is important to note, however, that these age effects were strongly influenced by the relatively short length of the retest intervals and possibly by the age–retest correlation. Because of this condition, the age effects primarily represent between-persons information and not much within-person change in age. The retest effects, on the other hand, are a completely within-person effect, and they could be estimated fairly precisely with multiple occasions and models with different growth functions.

The models used in our analyses rely on a convergence assumption (Bell, 1953) by which individuals of different age cohorts are followed over time. From this data structure with incomplete observations, one can model an overall trajectory for all persons as long as there is enough overlap or convergence in the trajectories of the different cohorts. This approach has been used (e.g., Duncan, Duncan, & Hops, 1996; McArdle & Bell, 2000) even under conditions of severe incomplete data (e.g., McArdle et al., 2002). In our analyses, this assumption, together with the wide age range, allowed us to estimate average age and retest effects as well as individual variation around those effects. The age effects represented an ongoing process from ages 40 to 74, although no individual in the study was actually measured across the entire range. The retest effects, in contrast, represented a process that occurred throughout the four measurement occasions. The multiple-group analyses indicated that the convergence assumption is reasonable for memory and space but does not hold for speed. For this variable, the effects associated with age and retest are not equal for all ages.

All models in this report were fitted to all the available data under an assumption of random dropout. This untested assumption is important in longitudinal studies. Attrition is likely to be selective in that individuals who continue are healthier, more motivated, and higher functioning than are those who discontinue their participation (Baltes & Mayer, 1999; Baltes, Reese, & Nesselroade, 1977; Hultsch et al., 1998; Schaie, 1996; Siegler & Botwinick, 1979). Although formal attrition analyses were not conducted, we repeated the main analyses with individuals who had complete data on all four measurement occasions ($N = 492$). Compared with the overall sample, those individuals who remained throughout the study had higher baseline scores but similar age effects for all variables. Practice effects were smaller for the memory and speed variables and larger for the space variable in this group, compared with the overall sample. Across all variables, the discrepancies in the estimates from the complete and incomplete samples were small. Future studies, however, could formally examine selective dropout as a function of each assessment, the number of occasions, and other selection conditions (e.g., Hedeker & Gibbons, 1999).

Future Research

The analyses reported here represent only a small set of possibilities for examining the effects of age and practice in studies of cognitive abilities. There are many ways in which these analyses can be expanded to accommodate more complex questions. For example, in our main analyses we used a linear function for age for all the models. Although a quadratic function fit slightly better, we decided to retain the linear trend for convenience, mainly to avoid overparameterization of the models. However, more complex age functions may be needed, especially for data with larger age ranges (McArdle et al., 2002).

A logical extension of the current analyses would be to apply multivariate models that consider different cognitive abilities simultaneously. One possible approach would involve examining how the changes in two or more variables may relate to each other (e.g., Goldstein, 1995; Willett & Sayer, 1994) and how they relate under different models of change (Ferrer & McArdle, 2002). One could also investigate whether the age and practice effects underlying the changes in each ability could be best described by changes in a higher order cognitive factor (as in McArdle, 1988; McArdle et al., 2002; McArdle & Woodcock, 1997). Another approach would be to examine the dynamics underlying changes in the cognitive abilities to identify time-lag sequences among the variables (Ferrer & McArdle, 2003; Hamagami & McArdle, 2001; McArdle, 2001; McArdle & Hamagami, 2001). In all these multivariate possibilities, the within-person changes may reflect both age and practice processes. If so, the changes in two variables can correlate because of age or practice, and these two components will bring different meaning to the correlation. Thus, modeling age and practice separately in multivariate approaches is important before interpreting correlations of change across variables.

Because of the short retest intervals in our study, the generalization of results about age and practice effects may be limited to studies with similar intervals. Other attempts should be made with data involving longer and more variable retest intervals, different cognitive abilities, and individuals of more different ages, as all these factors may influence age and retest differently (Cattell, 1957; McArdle & Woodcock, 1997). A potential benefit from such studies would be the establishment of guidelines about the number of measurement occasions and the optimal length of retest intervals needed to avoid contamination between practice and maturation, thus enabling researchers to better capture aging processes.

References

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

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Monterey, CA: Brooks/Cole.

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

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

Cattell, R. B. (1970). Separating endogenous, exogenous, ecogenic, and epogenic component curves in developmental data. *Developmental Psychology, 3*, 151–162.

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

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

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data with accelerated longitudinal designs. *Psychological Methods, 1*, 236–248.

Elias, M. F., Robins, M. A., & Elias, P. K. (1996). A 15-year longitudinal study of Halstead-Reitan Neuropsychological Test performance. *Journal of Gerontology Series B: Psychological Sciences and Social Sciences, 51*, 331–334.

Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 4*, 493–524.

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.

Hedeker, D., & Gibbons, R. D. (1999). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*, 74–78.

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

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

Hultsch, D. F., Hertzog, C., Small, B. J., McDonald-Miszczk, L., & Dixon, R. A. (1992). Short-term longitudinal change in cognitive performance in later life. *Psychology and Aging, 7*, 571–584.

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

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

Lowe, C., & Rabbitt, P. M. A. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36*, 1–8.

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

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

McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., 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., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38*, 115–142.

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

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

Rabbitt, P. M. A. (1993). Crystal quest: An examination of the concepts of "fluid" and "crystallized" intelligence as explanations for cognitive changes in old age. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention, selection, awareness, and control*. Oxford, England: Oxford University Press.

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford Press.

Schaie, K. W. (1988). Internal validity threats in studies of adult cognitive development. In M. L. Howe & C. J. Brainard (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 241–272). New York: Springer-Verlag.

Schaie, K. W. (1989). Perceptual speed in adulthood: Cross-sectional and longitudinal studies. *Psychology and Aging, 4*, 443–453.

Schaie, K. W. (1996). *Intellectual development in adulthood*. Cambridge, England: Cambridge University Press.

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

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

measurement and study duration. *Journal of Chronic Disease, 26*, 561–570.
Schwartz, B. S., Stewart, W. F., Bolla, K. I., Simon, D., Bandeen-Roche, K., Gordon, B., et al. (2000). Past adult lead exposure is associated with longitudinal decline in cognitive function. *Neurology, 55*, 1144–1150.
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.
Stewart, W. F., Schwartz, B. S., Simon, D., Bolla, K. I., Todd, A. C., & Links, J. M. (1999). Neurobehavioral function and tibial and chelatable lead levels in 543 former organolead workers. *Neurology, 52*, 1610–1617.
Thorndike, R. L. (1933). The effect of the interval between test and retest on the constancy of the IQ. *Journal of Educational Psychology, 24*, 543–549.
Vinsonhaler, J., & Meredith, W. (1966). A stochastic model for repeated testing. *Multivariate Behavioral Research, 4*, 461–477.
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.
Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17*, 179–193.
Wohlwill, J. F. (1970). The age variable in psychological research. *Psychological Review, 77*, 49–64.
Zelinski, E. M., & Burnight, K. P. (1997). Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. *Psychology and Aging, 12*, 503–513.
Zelinski, E. M., Gilewski, M. J., & Schaie, K. W. (1993). Individual differences in cross-section and 3-year longitudinal memory performance across the adult life span. *Psychology and Aging, 8*, 176–186.

Received May 21, 2003
Revision received November 10, 2003
Accepted January 5, 2004 ■