Developmental Psychology
2004, Vol. 40, No. 5, 813–822

Copyright 2004 by the American Psychological Association
0012-1649/04/$12.00 DOI: 10.1037/0012-1649.40.5.813

Estimating Retest Effects in Longitudinal Assessments of Cognitive Functioning in Adults Between 18 and 60 Years of Age

Timothy A. Salthouse
University of Virginia

David H. Schroeder
Johnson O'Connor Vocational Research Foundation

Emilio Ferrer
University of California, Davis

Several analyses were conducted on data from samples of adults between 18 and 58 years of age who completed the same cognitive tests after an interval ranging from less than 1 week to 35 years. Because the retest interval varied across individuals, it was possible to determine the length of time needed before the gains associated with a retest decreased to 0 and to obtain simultaneous estimates of the magnitude of effects associated with increased age and a prior assessment. The results indicated that for adults within this age range, 7 or more years were needed before positive retest effects were no longer detectable. Age effects in longitudinal comparisons could be interpreted in terms of large positive effects associated with a prior assessment and negative effects associated with age that were comparable in magnitude to those observed in cross-sectional comparisons.

For decades, a discrepancy has been recognized between the age trends in measures of cognitive functioning obtained from cross-sectional comparisons, which frequently reveal linear declines beginning when adults are in their early 20s, and the general absence of age-related declines before about age 60 in longitudinal comparisons. These different patterns were apparent in the earliest studies of aging and cognition based on the first group-administered test of cognitive ability for adults, the Army Alpha. To illustrate, linear declines in the total Army Alpha score from about age 20 to age 60 in cross-sectional comparisons were reported by Yerkes (1921) and Jones and Conrad (1933). However, Thorndike, Bregman, Tilton, and Woodyard (1928) reported that when adults between 16 and 40 years of age were tested again after an interval of 5–9 years, their second scores were higher, indicating positive longitudinal change. Recent data sets with longitudinal comparisons involving adults under the age of 60 also have revealed age-related declines in cross-sectional comparisons but higher performance on a second assessment, even when it occurred after an interval of 7 years (e.g., Huppert & Whittington, 1993; Schaie, 1996).

Because there are dramatically different implications for possible intervention and prevention of age-related cognitive decline if it starts when people are in their 20s rather than when they are in their 60s or later, it is important to understand the basis for the discrepancy between cross-sectional and longitudinal age trends. Moreover, depending on how the discrepancy is ultimately explained, it may also be relevant to the interpretation of relations between longitudinal change scores in samples of older adults because processes that have been underway for many years may reflect a different phenomenon than processes that are just beginning.

Three factors are commonly mentioned as potential contributors to different age trends in cross-sectional and longitudinal comparisons: historical shifts in absolute level of performance (e.g., cohort effects), selective attrition, and retest effects. However, there is still uncertainty about the duration, magnitude, and impact of each of these influences. For example, two quite different scenarios could be envisioned if aspects of the physical and cultural environment are changing in ways that lead to higher levels of cognitive performance as a function of historical time (i.e., the Flynn effect; Neisser, 1998; see also Tuddenham, 1948). One scenario is that susceptibility to these environmental changes diminishes with increased age, such that successive generations achieve higher asymptotic levels of cognitive functioning but then do not change much until late adulthood. An effect of this type would be somewhat analogous to the phenomenon of obsolescence and could result in serious distortions of age-related effects in cross-sectional comparisons. An alternative scenario is that the environmental change effects influence the cognitive functioning of people at all ages, perhaps in a manner analogous to the effect of inflation on salaries. In this scenario, it is the condition at the time of assessment that is critical, and thus comparisons across

Timothy A. Salthouse, Department of Psychology, University of Virginia; David H. Schroeder, Johnson O'Connor Vocational Research Foundation, Chicago, Illinois; Emilio Ferrer, Department of Psychology, University of California, Davis.

We acknowledge the institutional support of the Johnson O'Connor Research Foundation in this project, particularly for the collection of the data. We also recognize the contribution of Janine K. Bethscheider, former consultant for the Johnson O'Connor Research Foundation, who prepared the data for analysis and performed an earlier set of analyses.

Correspondence concerning this article should be addressed to Timothy A. Salthouse, Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22904-4400 or to David H. Schroeder, Johnson O'Connor Research Foundation, 161 East Erie Street, Room 304, Chicago, IL 60611. E-mail: salthouse@virginia.edu or research@jocrf.org

different points in time (i.e., longitudinal contrasts) could be misleading about the relations between age and cognitive functioning, as would likely be the case if one were to attempt to characterize the relations between age and salary from data collected at different points in time without adjusting for effects of inflation. It does not appear possible on the basis of available evidence to determine whether the obsolescence or the inflation analogy of historical changes in level of cognitive functioning is more appropriate, but it is important to realize that, depending on the assumptions, negative age-related effects could be either overestimated by cross-sectional comparisons or underestimated by longitudinal comparisons.

Selective attrition refers to the fact that people who return for subsequent assessments typically perform at a higher level on the initial assessment than people who do not return. Although this means that the samples in longitudinal studies are likely to be positively biased in comparison with the samples in cross-sectional studies, it does not necessarily follow that this would lead to different patterns of age relations when comparisons are restricted to individuals with data at each measurement occasion. The relative age trends in the two types of comparisons would only be affected if there is a relation between the initial level of functioning and the direction or rate of change, such that those at higher initial levels change in a different manner than those at lower initial levels. The question of whether age is kinder to the initially more able has been addressed in a number of studies, including three in which that phrase was used in the title of the article (i.e., Christensen & Hendersen, 1991; Deary, MacLennan, & Starr, 1999; Owens, 1959). However, results from these and other studies (e.g., Christensen et al., 2001; Mortensen & Hogh, 2001; Schaie, 1996; Sliwinski & Buschke, 1999) have suggested that the age trends in cognitive functioning are parallel across different initial levels. The relation between initial level and subsequent change warrants further investigation, particularly with long-term longitudinal studies, but on the basis of the available evidence, it does not appear that selective attrition, per se, is a major factor contributing to different age trends in cognitive functioning across cross-sectional and longitudinal comparisons.

Influences associated with a retest can be inferred to exist if individuals with prior testing experience perform better than otherwise comparable individuals of the same age who were tested at the same time but without previous testing experience. Although one might assume that estimates of retest effects could be obtained by contrasting the performance of individuals with and without a prior assessment, the situation is complicated by the phenomenon of selective attrition. That is, as Schaie (1988) noted, people with prior experience are not necessarily comparable with those without prior experience because the people who return for a second assessment often have a higher level of initial functioning than those who do not return. As a consequence, at least a portion of the difference in performance between those tested at a second occasion and those tested at the same time but for their first occasion may be attributable to a higher initial level of functioning among people who return for a second assessment. Better estimates of retest effects might therefore be obtained by comparing the performance of the same individuals on successive occasions.

Two aspects of retest effects can be distinguished: amount of gain across a short interval and the rate at which whatever gain is achieved decreases as a function of time. Results from comparisons used to compute test–retest reliability in standardized tests are relevant to the first aspect. At least with cognitive measures that are sensitive to age in cross-sectional comparisons, performance on the second assessment is typically higher than that on the first assessment. This phenomenon is illustrated in Table 1, which summarizes retest gains in several standardized tests in units of the standard deviation of the first measurement (i.e., Time 1 standard deviation [T1 SD] units). Inspection of the table reveals that the retest gains over a period of about 1 month averaged slightly less than 0.50 T1 SD units. It also appears that there is a tendency for the retest gains to be smaller with increased age, although it was not possible to evaluate the statistical significance of this trend.

Similar results have been reported in other studies with retest intervals up to 1 year. For example, Dikmen, Heaton, Grant, and Temkin (1999) found retest gains in performance IQ of 0.46 T1 SD units in a sample of adults with an average age of 34 and a retest interval of 9 months, and Snow, Tierney, Zorzitto, Fisher, and Reid (1989) reported an average retest gain of 0.11 T1 SD units in a sample of adults with an average age of 67 and a 1-year retest interval. Other studies with adults from a range of ages and retest intervals of 1 year or less have also reported retest gains of between 0.10 and 0.60 T1 SD units across a variety of variables (e.g., Benedict & Zgaljardic, 1998; Lowe & Rabbitt, 1998; Mitrushina & Satz, 1991; Salinsky, Storzbach, Dodrill, & Binder, 2001; Uchiyama et al., 1995; see review in Benedict & Zgaljardic, 1998). Furthermore, although the magnitude is typically smaller, retest gains have been found even when different materials or forms are used in the retest (e.g., Benedict & Zgaljardic, 1998; Dikmen et al., 1999; Watson, Pasteur, Healy, & Hughes, 1994; also see Rabbitt, Diggle, Smith, Holland, & McInnes, 2001).

On the basis of the results just described, it is possible that both positive retest influences and negative age influences are operating simultaneously in longitudinal comparisons, particularly among young adults for whom the positive retest effects may be large. However, little is currently known about the duration of retest

Table 1
Short-Term Retest Gains (in Time 1 Standard Deviation Units) From Standardized Tests

Source and age range	Mean retest interval (in days)	*N*	Gain
Wechsler (1981, Table II)			
25–34	14–49	71	0.57
45–54	14–35	48	0.60
Wechsler (1997, Tables 3.6–3.9)			
16–29	35	100	0.67
30–54	35	102	0.60
55–74	35	104	0.40
75–89	35	88	0.24
WASI (1999, Table 5.4)			
17–54	31	51	0.32
55–89	31	55	0.18
Kaufman & Kaufman (1993, Table 8.2)			
20–54	31	51	0.50
55–85+	31	54	0.43

Note. The variable for Wechsler (1981, 1997) and WASI (Wechsler Abbreviated Scale of Intelligence; 1999) was the Performance Intelligent Quotient; the variable for Kaufman and Kaufman (1993) was the composite measure of fluid cognitive abilities.

effects or of their magnitude relative to effects associated with increased age. Perhaps the simplest way to investigate the duration of retest effects is to compare the size of the difference between a given variable at the first (T1) and second (T2) measurement occasion as a function of the interval between occasions. A comparison of this type requires that people vary in the length of the interval between the two assessments (cf. McArdle & Woodcock, 1997), but to the extent that there is variability in the T1 − T2 interval, it should be possible to estimate both the magnitude of the retest effect and the time interval required for it to diminish to 0.

Another advantage of a design in which there is variability among people in the retest intervals is that one can simultaneously estimate the magnitude of age effects and of retest effects. That is, in a typical two-occasion longitudinal study, it is difficult to decompose observed change scores into gains associated with the retest and possible negative effects associated with increased age, because when everyone has the same retest interval there is a confounding between the increment in age and the increment in occasion number (or amount of prior experience). However, when people vary in the interval between the first and second assessments, the increase in age is no longer perfectly correlated with the interval between assessments, and it becomes possible to distinguish the two components.

McArdle, Ferrer-Caja, Hamagami, and Woodcock (2002) recently reported analyses of data collected with a variable retest interval design in which the intervals from the first to the second occasion ranged from a few months to 10 years. In addition to providing estimates of the age relations from cross-sectional and longitudinal comparisons, the analyses yielded separate estimates of the effects associated with age and the effects associated with retest. To illustrate, for the fluid cognitive ability measure, a simple regression analysis of the first occasion cross-sectional data for adults between 20 and 49 years of age yielded a slope of −.362, indicating that each year was associated with a decrease of .362 units.[1] In contrast, the estimate of observed longitudinal change, based on the difference in score from the first to the second occasion divided by the difference in age from the first to the second occasion, was 2.0 units. A particularly informative aspect of the McArdle et al. (2002) study was the use of mixed effects analyses, based on the following equation, to simultaneously derive estimates of the age and retest effects.

$$Y_{m,n} = Y_{0,n} + \mathrm{Age}_{m,n} \times \Delta Y_n + R_m \times \Delta R_n + e_{m,n} \qquad (1)$$

where $Y_{m,n}$ is the observed score of person n at measurement m, $Y_{0,n}$ is the latent initial score of person n, $\mathrm{Age}_{m,n}$ is the observed age of person n at measurement m, ΔY_n is the latent age slope of person n, R_m is the retest effect at measurement m, ΔR_n is the latent retest slope of person n, and $e_{m,n}$ is the latent error score of person n at measurement m. The subscript n refers to the random components of the model in that they represent parameters that can vary across people. It is important to note that in this type of analysis, the age variable is scaled in years, but the retest variable is scaled in terms of number of prior assessments (i.e., 0 for T1 and 1 for T2).

The estimate of the average age slope (i.e., ΔY) for adults between 20 and 49 in this analysis was −.389, and the estimate of the average practice slope (i.e., ΔR) was 5.85. The estimated annual decrease associated with increased age in this analysis was therefore nearly the same as that found in the cross-sectional comparison (i.e., −.389 vs. −.362), but the effect of a single retest was over 15 times greater than the estimated annual decline. The combination of the two effects within this age range would have resulted in the negative age-related effects being completely obscured by the large positive retest effect had it not been possible to estimate both effects in the analysis.

Similar types of analyses attempting to distinguish age and retest effects have been reported in other recent studies involving older adults, and in each case the estimated retest effects were large relative to the estimated age effects (e.g., Ferrer, Salthouse, Stewart, & Schwartz, 2004; Rabbitt et al. 2001; Wilson et al., 2002). Furthermore, each group of researchers reached very similar conclusions. For example, Rabbitt et al. (2001) suggested that "all studies, which have not taken practice effects into account, have seriously underestimated the rates of cognitive decline in old age" (p. 540); Wilson et al. (2002) claimed that, "It is likely, therefore, that the observed levels of decline in cognitive test performance underestimate the actual levels of declines in the abilities being measured" (p. 190); and Ferrer et al. (2004) noted that "ignoring the retest process in the model tends to underestimate the age effects, with more severe bias occurring when retest effects are larger" (p. 255).

There were two goals of the present study. The first was to compare the magnitude of the gains from the first to the second occasion with the magnitude of the age-related differences derived from cross-sectional comparisons and to estimate the interval between successive assessments needed for the gain associated with a retest to decrease to zero. The second goal of the study was to apply the same type of mixed effects models used by McArdle et al. (2002) to six variables previously established to have significant negative relations to age in cross-sectional comparisons of adults between 20 and 50 years of age (i.e., Schroeder & Salthouse, 2004). Of particular interest was whether we would be able to replicate the findings of nearly comparable cross-sectional and longitudinal estimates of age-related declines in cognitive functioning among adults under the age of 60 after taking retest effects into consideration.

The data available for analysis in the current study consisted of the cognitive scores on an initial and subsequent assessment occasion for samples of 120–284 adults between 18 and 58 years of age. Although participants were only assessed on two occasions, thus restricting analyses of within-person changes to linear relations, the interval between the first and the second occasion varied from less than a week to 35 years. This feature of the data allowed the longitudinal changes to be decomposed into a portion attributable to aging and to a portion attributable to the retest.

Method

The data were collected from individuals ($N > 15{,}000$) who were clients at various offices of the Johnson O'Connor Research Foundation across the United States. All of them paid, either directly or through their employer, for an initial vocational assessment. Some of the individuals (see Table 2 for sample sizes) returned for further consultation, and if so, they were asked to retake one of the tests. Because any given individual repeated only one test, the samples for the different tests did not overlap. The retest intervals ranged from a few days to 35 years.

The present study focuses on six variables that had significant negative age correlations in a previous cross-sectional comparison (Schroeder &

[1] Some of the results described here are based on new analyses of the data for adults between 20 and 49 years of age.

Table 2
Descriptive Characteristics (Means and Standard Deviations) of Samples at the First Assessment

	Age range			
Variable and characteristic	18–29	30–39	40–58	All
Associative Learning				
N	87	42	27	156
Mean age	22.5 (3.3)	34.6 (2.8)	44.9 (3.6)	29.6 (9.3)
Mean retest interval	4.9 (3.8)	3.5 (3.5)	2.5 (2.5)	4.1 (3.6)
Wiggly Block				
N	65	27	27	120
Mean age	23.1 (3.3)	34.5 (3.1)	44.8 (3.8)	30.6 (9.5)
Mean retest interval	2.5 (2.2)	1.8 (1.8)	1.6 (1.8)	2.1 (2.1)
Recognition Memory				
N	130	41	25	197
Mean age	23.1 (3.6)	35.1 (2.3)	46.0 (4.0)	28.7 (8.9)
Mean retest interval	5.8 (4.8)	3.1 (3.7)	4.2 (3.9)	5.0 (4.6)
Number Memory				
N	123	33	25	181
Mean age	22.1 (3.4)	34.8 (2.8)	46.0 (4.6)	27.7 (9.4)
Mean retest interval	5.5 (6.1)	3.7 (4.4)	2.9 (2.8)	4.8 (5.5)
Memory for Design				
N	110	37	25	172
Mean age	22.4 (3.3)	34.9 (3.0)	45.4 (3.4)	28.4 (9.2)
Mean retest interval	5.7 (5.2)	2.6 (3.1)	2.4 (2.6)	4.6 (4.7)
Inductive Reasoning				
N	187	71	26	284
Mean age	22.8 (3.7)	34.8 (2.8)	47.1 (5.0)	28.0 (8.7)
Mean retest interval	3.2 (3.0)	3.1 (3.3)	1.6 (2.1)	3.1 (3.0)

Note. Age and retest interval are expressed in years. Numbers in parentheses are standard deviations.

Salthouse, 2004). The Associative Learning Test involved paired-associates learning and memory for verbal material (i.e., words and nonsense syllables). The Wiggly Block Test required manipulation of three-dimensional blocks containing curved surfaces to assemble a cube. The Recognition Memory Test assessed recognition memory for the identity and position of objects from a set of pictures. The Number Memory Test assessed recall memory for sets of numbers, and the Memory for Designs Test assessed accuracy in reproducing line patterns. Finally, the Inductive Reasoning Test measured quickness in selecting which pictures, from among a set of six, had something in common.

Characteristics of the retest sample for each variable, including the age and retest interval, are summarized in Table 2. Note that all of the participants were under the age of 60 years, with a large proportion under the age of 30. It is also important to recognize that unlike traditional longitudinal designs in which everyone is retested at nearly the same interval, there is considerable variability in the retest intervals for each variable. In fact, examination of Table 2 reveals that in most cases the standard deviation of the retest interval was nearly as large as the mean retest interval. Correlations between age at the first assessment and the retest intervals ranged from −.15 (for Inductive Reasoning) to −.32 (for Memory for Designs), indicating that, on the average, the retest interval was somewhat shorter for older individuals. For none of the variables was there a significant correlation between the score at the first occasion and the length of the interval until the second occasion (i.e., range from −.10 to .08, *Mdn* = −.01).

Results

The dotted lines in the six panels of Figure 1 portray the cross-sectional age relations for the total sample of individuals assessed on these variables, and for the retest sample at the first (T1) and second (T2) measurement occasions. The solid line represents the average change across the retest interval for three age groups. For most of the variables, the absolute level of performance and the relations with age were quite similar in the total sample and in the retest sample at the first (T1) occasion. The primary exceptions were that adults in the 18-to-29 age group in the retest samples for the Associative Learning and Wiggly Block variables had higher average levels of performance than did the total sample. Although these individuals may have differed from the total sample in certain characteristics, for the most part the individuals who returned for a retest appeared comparable to the total sample of individuals who were assessed with these tests.

In each of the panels, all three cross-sectional age relations were negative, which confirms that for these cognitive measures, increased age in the range from 18 to 60 years is associated with lower levels of performance. It is also apparent in each of the panels that the level of performance at the second occasion (T2) was much higher than that at the first occasion (T1) and that the solid lines representing average within-person changes all had positive slopes. The results portrayed in Figure 1 therefore replicate the pattern of pronounced age-related declines in cross-sectional comparisons (dotted lines) and improvements in longitudinal comparisons (solid lines) for adults under the age of 60.

The initial analyses of the data examined linear and quadratic effects of age on the score at T1 and of the retest interval on the change in score from T1 to T2. Nearly all of the linear effects were significant (i.e., median R^2 = .042 for age and .058 for retest interval), but none of the quadratic effects were significant (i.e., median R^2 = .004 for both age and retest interval). Only linear relations of age and retest interval were therefore included in the subsequent analyses. To facilitate across-variable comparisons, all of the variables were expressed in a common scale by converting the scores to standard deviations of the first measurement occasion.

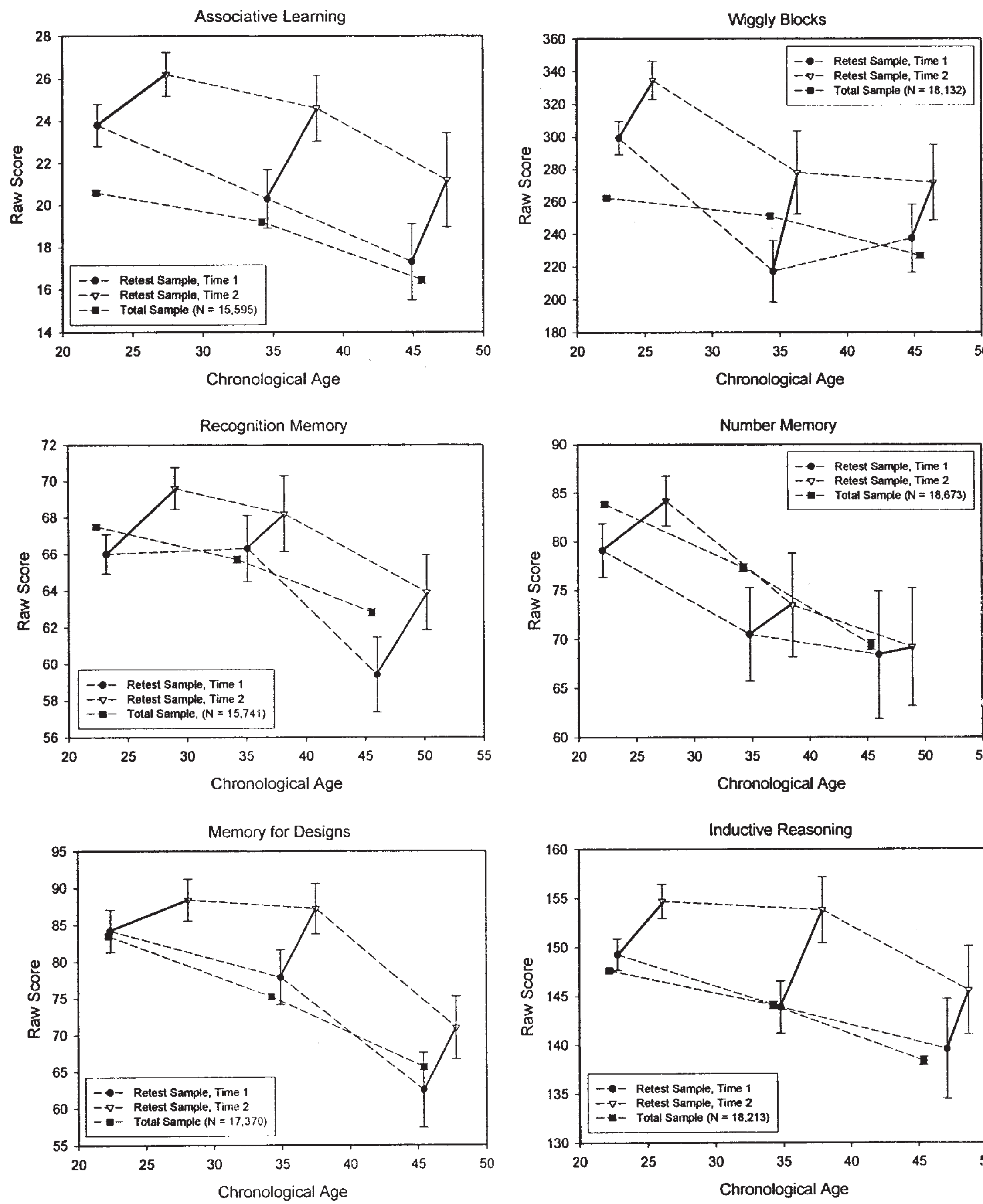

Figure 1. Cross-sectional (dotted lines) and longitudinal (solid lines) age relations in the total and retest sample at the first (Time 1) and second (Time 2) measurement occasions. Bars above and below each point are standard errors.

The observed change from T1 to T2 was used as the dependent variable in the analyses designed to investigate the persistence of the retest effects. The T2 − T1 change scores were plotted as a function of the retest interval in Figure 2. It can be seen that for all of the variables the values with a retest interval of less than 1 year were about 0.5 SD units, which is comparable to the values summarized in Table 1 for adults in this age range.

The T2 − T1 change scores for each variable were analyzed in hierarchical regression analyses with age, retest interval, and the cross-product (Age × Retest Interval) interaction terms entered successively. None of the age effects were significant (i.e., $R^2 <$ .018), but all of the effects of retest interval were significantly different from zero (i.e., $R^2 =$ .023 to .142). Only one interaction was statistically significant (i.e., $R^2 =$ .045, for the Number Memory variable), and it was attributable to a smaller effect of retest interval with increased age. These results suggest that within this age range the magnitude of the gain in performance from the first to the second assessment was not related to age, nor did the

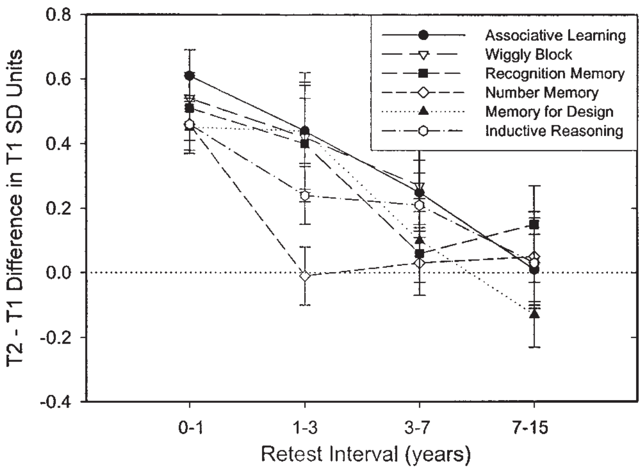

Figure 2. Retest gains in six cognitive variables as a function of retest interval. Bars above and below each point are standard errors. T1 = first measurement occasion; T2 = second measurement occasion.

relation between the T1 to T2 interval and the retest gain vary as a function of age.

Two sets of regression equations were next computed relating the T2 − T1 difference to the retest interval to determine the interval required for the difference to reduce to zero. The first set of equations was based on data from all retest participants, and the second set was based on data for those participants with a retest interval equal to or greater than 1 year. The intercepts and slopes of these equations were then used to estimate the number of years until the retest gain was eliminated, corresponding to a T2 − T1 difference of zero. That is, the regression equation, $Y = a + b(X)$, where Y is the T2 − T1 performance difference, and X is the retest interval, was rearranged to determine the X value corresponding to a Y value of zero.

Results of these regression analyses are summarized in Table 3. Inspection of the table reveals that for every variable, at least 7 years were needed for the gain to diminish to zero. Furthermore, with the exception of the Number Memory variable, the results were similar with and without data from participants with retest intervals of less than 1 year. The anomalous results with the Number Memory variable are likely attributable to the near-zero retest gains for intervals greater than 1 year, apparent in Figure 2.

The regression equations based on the data from all retest participants were then used to estimate the expected retest gain across a 3-year interval, and these estimated retest gains were compared with the expected cross-sectional difference across a 3-year age range based on the data from the total sample of adults and on the data from the retest sample at the first occasion. These three values are displayed in the right-most columns of Table 3. It can be seen that the estimated retest gains across a 3-year retest interval ranged from 0.20 to 0.41 T1 SD units, whereas the average cross-sectional difference was between −0.04 and −0.10 T1 SD units. If the cross-sectional age relations are assumed to approximate the true maturational age relations, these results imply that for adults between about 18 and 60 years of age, the positive retest effects across a 3-year interval are between 4 and 10 times larger than the negative maturational effects.

The next set of analyses compared estimates of the age relations from conventional cross-sectional and longitudinal comparisons. That is, estimates of the cross-sectional age difference (based on the slope of the regression equation relating score to age) and of the longitudinal age change (derived by dividing the difference in score from T2 to T1 by the difference in age from T2 to T1) were computed for each variable. These estimates are presented in the first through the fourth columns of Table 4.

The values in the first and second columns represent the comparison of the total sample and the retest sample at T1. As noted earlier, for a few variables, the retest sample differed from the total sample either in the initial level (e.g., Associative Learning, Wiggly Blocks) or in the age relation (e.g., Wiggly Blocks). The higher intercepts in the third column (T2) compared with those in the second column (T1) reflect the higher performance at the second occasion, but the generally similar age slopes suggest that the gains associated with a retest were comparable at every age. The age slopes range from −.013 to −.034, which correspond to differences of between 0.65 and 1.7 standard deviation units over 50 years. These latter values are in the range of the estimates obtained in cross-sectional studies with samples of 200 or more adults across a wide age range (e.g., Park et al., 2002; Schaie, 1996; Salthouse, 1993, 1994; Salthouse & Czaja, 2000; Salthouse & Ferrer-Caja, 2003; Salthouse, Hambrick, & McGuthry, 1998; Verhaeghen & Salthouse, 1997).

The values in the fourth column represent the average within-person change per year based on the longitudinal comparison (i.e., difference in score divided by difference in age). These values are not only in the opposite direction of the estimated cross-sectional age effects but are between 13 and 57 times greater in magnitude. To illustrate, the average estimated age relation for the Associative Learning variable in the cross-sectional comparisons was −.028 units per year, whereas in the longitudinal comparison it was .980 units per year.

Values in the fifth and sixth columns of Table 4 were obtained using Equation 1 after first centering the age variable by subtracting the youngest age, 18 years, from all ages.[2] Two analyses were conducted with each dependent variable. The first included age (scaled in years), retest (coded as 0 for *no* and 1 for *yes*), and the interaction of age and retest as fixed effects, with random effects for the intercept (initial level of Y) and for age (estimated change in Y as a function of age). None of the interactions of age and retest were significant with any of the variables (i.e., all p values greater than .13), and thus the interaction term was dropped from the second analysis. In addition, the covariation between the age and retest variables was set to zero, and random coefficients were estimated for the intercept, age, and retest effects (as in McArdle

[2] Parallel analyses were also conducted when the analyses were repeated after centering the age variable for each individual on the basis of the average of his or her ages across the two measurement occasions. Because the sum of the two person-centered age variables equals the retest interval, it is not surprising that the correlations between the person-centered age variable and the binary coded (0 for none, 1 for one) retest variable ranged from .66 to .75. Nevertheless, the same pattern of small but significant negative age effects and large positive retest effects was found in these analyses. In particular, the fixed effects coefficients for the age and retest effects, respectively, were −.068 and .620 for Associative Learning, −.071 and .573 for Wiggly Block, −.037 and .447 for Recognition Memory, −.043 and .313 for Number Memory, −.066 and .499 for Memory for Designs, and −.038 and .409 for Inductive Reasoning.

Table 3
Estimated Duration of Retest Effects and the Magnitude of Retest Gains and of Cross-Sectional Differences Across a 3-Year Interval

Variable	Years until retest gain = 0: All participants	Years until retest gain = 0: Excluding participants with retest intervals < 1 year	T1 SD units across 3 years: Retest gain	T1 SD units across 3 years: Cross-sectional age difference: Total	T1 SD units across 3 years: Cross-sectional age difference: Retest
Associative Learning	9.0	9.2	0.41	−0.05	−0.08
Wiggly Block	7.3	7.3	0.35	−0.04	−0.10
Recognition Memory	11.9	12.8	0.36	−0.05	−0.05
Number Memory	9.0	0.4	0.20	−0.06	−0.05
Memory for Design	7.9	8.0	0.30	−0.08	−0.09
Inductive Reasoning	9.9	13.4	0.29	−0.05	−0.04

Note. T1 SD = Time 1 standard deviation.

et al., 2002).[3] Parameter estimates from this second set of analyses are presented in the rightmost columns of Table 4.

There are two important results from these analyses. The first is that the retest effects were all positive and moderately large, with a range of between 9 and 20 times greater than the estimated annual decline associated with age. The second important result is that after taking effects associated with the retest into consideration, the estimates of the age effects were similar in direction to, and if anything larger in magnitude than, the estimates derived from the cross-sectional comparisons. For example, the estimated annual age effect for the Associative Learning variable in the cross-sectional analysis at T1 was −.027 standard deviation units, and it was −.037 standard deviation units in the analysis in which the age and retest effects were estimated simultaneously.

Because the estimates of the age effects in Equation 1 are based on both cross-sectional (between-persons) and longitudinal (within-person) information, and because the range of cross-sectional age is considerably larger than the range of longitudinal retest intervals (cf. Table 2), the estimates of the age effects may have been dominated by the cross-sectional age relations. In order to investigate this possibility, the analyses were repeated on the subsample of adults under the age of 30. This group was selected because the sample sizes were relatively large, and the standard deviations for the age and retest interval variables were similar in magnitude. Results from these analyses (not displayed in tables) closely resembled those reported in Table 4. To illustrate, for the six variables listed successively in Table 4, the estimates of the age effects for adults between 18 and 29 years of age were −.039, −.050, −.042, −.032, −.059, and −.010, respectively, and the estimates of the retest effects were .458, .482, .539, .318, .482, and .276, respectively. Not only are these values close in absolute magnitude to those based on the complete sample (cf. Table 4), but the ratios of the retest effects to the age effects were also similar (i.e., range = 8–27, *Mdn* = 10.3). It therefore does not appear that the estimates of the age effects in Table 3 are artifacts of the larger range of between-persons (cross-sectional) ages than of within-person (longitudinal) retest intervals.

The mixed effects models also yielded estimates of the variability exhibited by the individuals around the average fixed effects. These estimates, presented in the column labeled *Random* in Table 4, indicated that people varied considerably in the intercept or initial level of *Y* but that there was very little across-individual variability in the magnitude of the age effects (i.e., ΔY). The absolute values of the estimates of variability for the retest effects (i.e., ΔR) were moderately large, but so were the standard errors, and none of the estimates were significantly different from zero. These results therefore suggest that although people differ substantially in the initial level of performance, at least within this range of age and retest intervals, they were relatively similar in the degree to which their performance changed as a function of age or retest.

Discussion

Two different types of analyses converged on the conclusion that, at least in adults under the age of 60, the increases in performance associated with a prior assessment are quite large relative to the decreases associated with increased age. One analysis simply compared estimates of the magnitude of the gain from a first to a second assessment with estimates of the magnitude of age differences based on cross-sectional comparisons. Because the interval between the assessments varied from a few days to over 35 years, it was also possible to estimate the interval required for the retest effects to be eliminated. The analyses revealed that between 7 and 13 years must elapse before the advantage of the prior assessment was eliminated for these variables. The second analysis was based on a statistical model of age and retest effects that capitalized on the fact that the interval between the first and the second measurement occasion varied across participants. Results from both sets of analyses indicate that, for adults under the age of 60 and retest intervals less than about 7 years, retest effects may be greater than age effects by as much as a factor of 15, and thus the observed longitudinal change will frequently be positive despite the presence of age-related decline. These findings therefore suggest that a major reason for the discrepancy between age trends observed in cross-sectional and longitudinal comparisons is that negative age effects in longitudinal comparisons are completely obscured by positive retest effects.

Although the two analytical procedures were consistent in yielding estimates of small negative age effects and large positive retest effects, it is important to recognize that they involved somewhat

[3] Because there were only two measurement occasions, the covariance of the random age effects and random retest effects had to be fixed to zero for the models to converge.

Table 4
Estimates of Age and Retest Effects From Cross-Sectional, Longitudinal, and Mixed-Effects Analyses

Variable and parameter	Cross-sectional Total	Cross-sectional T1	Cross-sectional T2	Longitudinal ($\Delta Y/\Delta$age)	Mixed-effects Fixed	Mixed-effects Random
Associative Learning						
Intercept	.268*	.796*	1.323*		.419*	.826*
Age	−.017*	−.027*	−.029*	0.980*	−.037*	.001
Retest					.493*	.121
Wiggly Blocks						
Intercept	.216*	1.013*	1.544*		.496*	.520
Age	−.013*	−.033*	−.034*	1.435*	−.041*	.001
Retest					.508*	.242
Recognition Memory						
Intercept	.578*	.474	1.049*		.240	.792*
Age	−.018*	−.017	−.023*	0.533*	−.022*	.000
Retest					.388*	.187
Number Memory						
Intercept	.659*	.443	1.249*		.276*	.914*
Age	−.019*	−.016	−.034*	0.817*	−.027*	.001
Retest					.260*	.000
Memory for Designs						
Intercept	.755*	.840*	1.334*		.419*	.966*
Age	−.027*	−.030*	−.034*	0.411*	−.040*	.001
Retest					.383*	.000
Inductive Reasoning						
Intercept	.348*	.353	.822*		.168	.570*
Age	−.016*	−.013	−.017	0.860*	−.017*	.000
Retest					.345*	.134

Note. The estimates for the mixed-effects models were obtained with maximum likelihood procedures with SAS Proc Mixed.
* $p < .01$.

different sets of assumptions. For example, the critical assumption underlying the comparison of T2 and T1 performance to estimate retest effects is that there were minimal positive changes across that interval in factors unrelated to retest. If there were increases in ability from the first to the second assessment or if the environment changed in ways that contributed to increases in performance on these tasks, then the retest estimates would be overestimated in this analysis. (Conversely, if there were decreases in ability or negative changes in relevant aspects of the environment across the retest interval, then the inferred retest gains would be underestimated with this procedure.) The mixed effects analyses took advantage of the fact that the retest interval varied across individuals to separate influences associated with age and retest, and it is based on the assumption that it is meaningful to combine within-person and between-persons age effects in order to distinguish them from retest effects. This assumption can be questioned, particularly because of the difficulty of determining the relative contribution of the between-persons and within-person influences on the overall age estimate. However, the critical estimate for the current purposes concerns the magnitude of the retest effect, and that estimate is only partially influenced by the composition of the estimated age effect.

The similar patterns of results with different sets of assumptions is reassuring, but alternative analytical methods should continue to be explored because of the inherent difficulty of disentangling between-persons age differences, within-person age changes, and within-person retest changes in two-occasion longitudinal data. One possibility worth considering in future longitudinal studies is to combine variable retest intervals with more than two occasions to allow more powerful decomposition of the three types of influences.

The discrepancy between the estimates of the age-related effects in cross-sectional and longitudinal comparisons might not be expected to be as large among older adults if, as the short-term retest results summarized in Table 1 suggest, the retest effects are smaller with increased age. Indeed, Wilson et al. (2002) reported age relations of similar magnitude in cross-sectional and longitudinal comparisons across seven composite variables in adults with an average age of 76, and Sliwinski and Buschke (1999) actually reported larger negative age relations in longitudinal comparisons than in cross-sectional comparisons for 12 of 13 variables for adults between 66 and 92 years of age.

Although similar retest effects were observed in each variable in the analyses reported here, it is also possible that the magnitude of the retest effects may vary across variables in samples with older participants. In fact, Ferrer et al. (2004) recently found that this was the case in which the same type of analytical model used here was applied to data from a sample of adults between 40 and 72 years of age with annual retests across a period of 4 years. Three composite variables were analyzed, and large retest effects were found for a memory variable, small retest effects for a speed variable, and intermediate effects for a measure of spatial ability.

At least four types of influences may contribute to the retest gains that occur across successive measurement occasions. The most obvious are test-specific factors such as remembering the answers to particular items. General familiarity with the testing situation, perhaps accompanied by a reduction in anxiety, may also contribute to improved performance, even when different forms are used in each occasion. Another possibility that could contribute to better performance on subsequent occasions is an increase in the relevant ability or skills during the interval between measurement occasions. Finally, in some circumstances, a portion of the gain from the first to the second measurement occasion might be attributable to changes that occur in the environment in which the individual lives. An influence of this type is probably most plausible for general information or vocabulary tests. For example, a question about the capital of a relatively unfamiliar country might be easier to answer after there has been considerable publicity about the country because of a war, epidemic, or famine. Because the retest effects appear to persist for as long as 13 years, it would clearly be desirable to disentangle the various types of retest influences and to determine the relative time course of each. Although this does not appear possible with the current design and analytical procedures, it may be feasible in the future by systematically combining variable retest intervals with systematic manipulations of the similarity of the test materials across occasions and by adding time-lag comparisons in which different people of the same ages are tested for the first time with each administration of a retest. Comparison of retest gains across different test versions should be informative about the role of test-specific factors, and time-lag comparisons would provide an estimate of gains attributable to cultural or environmental changes.

To reiterate, the results of the analyses reported here suggest that a substantial proportion of the discrepancy between estimates of age–cognition relations from cross-sectional and longitudinal comparisons that involve adults under age 60 is not because cross-sectional results overestimate the true age relations but because the longitudinal comparisons, at least with retest intervals of less than 10 years, incorporate large positive retest effects that may overwhelm the negative age effects. A conclusion of this type obviously challenges conventional assumptions that little or no cognitive change occurs before about age 60. However, it is consistent with findings of nearly linear cross-sectional declines beginning in the 20s in certain neurophysiological parameters that might serve as substrates of cognitive functioning, such as the density of dopamine receptors (e.g., Backman et al., 2000; Kaasinen et al., 2000; Volkow et al., 1998, 2000), relative concentration of brain metabolites (e.g., Kadota, Horinouchi, & Kuroda, 2001), measures of the integrity of myelin (e.g., Abe et al., 2002; Nusbaum, Tang, Buchsbaum, Wei, & Atlas, 2001; O'Sullivan, Jones, Summers, Morris, Williams, & Markus, 2001), volume of white matter (Guttmann et al., 1998), volume of gray matter (Bartzokis et al., 2001; Good et al., 2001), density of gray matter (Sowell et al., 2003), and the volumes of various brain regions (e.g., Raz, 2000; Tisserand, Visser, van Boxtel, & Jolles, 2000).

The suggestion that age-related cognitive decline begins when people are in their 20s also has implications for the timing of potential interventions. That is, if the goal is prevention, instead of remediation, it may be necessary to target the period of young adulthood. Although interventions at any age may be effective in increasing the level of performance in specific tasks, they may not be very successful in altering the relation between age and performance unless they occur when the declines are first occurring.

The findings of significant negative relations between age and measures of cognitive functioning before age 60 in both cross-sectional and retest-adjusted longitudinal analyses also lead to questions about how results involving correlations among longitudinal change scores for different cognitive variables in older adults should be interpreted if the change in both variables actually started decades earlier. If it is really the case that the phenomenon of cognitive aging starts when adults are in their 20s or 30s, then attempting to identify causes of the phenomenon by examining correlations among change in adults in their 60s or later may be analogous to attempting to infer which rocks precipitated an avalanche by examining their rate of progress half way down the mountain. Of course, it is possible that factors in childhood or earlier serve to determine when, and by how much, cognitive functioning is affected by increased age in adulthood. However, results such as those described here suggest that the phenomenon of age-related cognitive decline can be detected relatively early in adulthood with both cross-sectional and longitudinal comparisons when the contribution of retest gains in longitudinal data is taken into consideration.

References

Abe, O., Aoki, S., Hayashi, N., Yamada, H., Kunimatsu, A., Mori, H., et al. (2002). Normal aging in the central nervous system: Quantitative MR diffusion-tensor analysis. *Neurobiology of Aging, 23,* 433–441.

Backman, L., Ginovart, N., Dixon, R. A., Wahlin, T. B. R., Wahlin, A., Halldin, C., & Farde, L. (2000). Age-related cognitive deficits mediated by changes in the striatal dopamine system. *American Journal of Psychiatry, 157,* 635–637.

Bartzokis, G., Beckson, M., Lu, P. H., Neuchterlein, K. H., Edwards, N., & Mintz, J. (2001). Age-related changes in frontal and temporal volumes in men: A magnetic resonance imaging study. *Archives of General Psychiatry, 58,* 461–465.

Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology, 20,* 339–352.

Christensen, H., & Hendersen, A. S. (1991). Is age kinder to the initially more able? A study of eminent scientists and academics. *Psychological Medicine, 21,* 935–946.

Christensen, H., Hofer, S. M., Mackinnon, A. J., Korten, A. E., Jorm, A. F., & Henderson, A. S. (2001). Age is no kinder to the better educated: Absence of an association investigated using latent growth techniques in a community sample. *Psychological Medicine, 31,* 15–28.

Deary, I. J., MacLennan, W. J., & Starr, J. M. (1999). Is age kinder to the initially more able? Differential ageing of a verbal ability in the Healthy Old People in Edinburgh Study. *Intelligence, 26,* 357–375.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test–retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society, 5,* 346–356.

Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging, 19,* 243–259.

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage, 14,* 21–36.

Guttmann, C. R. G., Jolesz, F. A., Kikinis, R., Killiany, R. J., Moss, M. B., Sandor, T., et al. (1998). White matter changes with normal aging. *Neurology, 50,* 972–978.

Huppert, F. A., & Whittington, J. E. (1993). Longitudinal changes in mental state and personality measures. In B. D. Cox, F. A. Huppert, & M. J. Whichelow (Eds.), *The Health and Lifestyle Survey: Seven years on* (pp. 133–172). Hants, England: Dartmouth Publishing.

Jones, H. E., & Conrad, H. (1933). The growth and decline of intelligence:

A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychological Monographs, 13,* 223–298.
Kaasinen, V., Vilkman, H., Hietala, J., Nagren, K., Helenius, H., Olsson, H., et al. (2000). Age-related dopamine D2/D3 receptor loss in extrastriatal regions of the human brain. *Neurobiology of Aging, 21,* 683–688.
Kadota, T., Horinouchi, T., & Kuroda, C. (2001). Development and aging of the cerebrum: Assessment with proton MR spectroscopy. *American Journal of Neuroradiology, 22,* 128–135.
Kaufman, A. S., & Kaufman, N. L. (1993). *Manual for the Kaufman Adolescent and Adult Intelligence Test (KAIT).* Circle Pines, MN: American Guidance Service.
Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36,* 915–923.
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., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods, 2,* 403–435.
Mitrushina, M., & Satz, P. (1991). Effect of repeated administration of a neuropsychological battery in the elderly. *Journal of Clinical Psychology, 47,* 790–801.
Mortensen, E. L., & Hogh, P. (2001). A gender difference in the association between APOE genotype and age-related cognitive decline. *Neurology, 57,* 89–95.
Neisser, U. (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.
Nusbaum, A. O., Tang, C. Y., Buchsbaum, M. S., Wei, T. C., & Atlas, S. W. (2001). Regional and global changes in cerebral diffusion with normal aging. *American Journal of Neuroradiology, 22,* 136–142.
O'Sullivan, M., Jones, D. K., Summers, P. E., Morris, R. G., Williams, S. G. R., & Markus, H. S. (2001). Evidence for cortical "disconnection" as a mechanism of age-related cognitive decline. *Neurology, 57,* 632–638.
Owens, W. A. (1959). Is age kinder to the initially more able? *Journal of Gerontology, 14,* 334–337.
Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17,* 299–320.
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.
Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 1–90). Mahwah, NJ: Erlbaum.
Salinsky, M. C., Storzbach, D., Dodrill, C. B., & Binder, L. M. (2001). Test–retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12–16-week period. *Journal of the International Neuropsychological Society, 7,* 597–605.
Salthouse, T. A. (1993). Speed mediation of adult age differences in cognition. *Developmental Psychology, 29,* 722–738.
Salthouse, T. A. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology, 30,* 240–259.
Salthouse, T. A., & Czaja, S. J. (2000). Structural constraints on process explanations in cognitive aging. *Psychology and Aging, 15,* 44–55.
Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging, 18,* 91–110.
Salthouse, T. A., Hambrick, D. Z., & McGuthry, K. E. (1998). Shared age-related influences on cognitive and non-cognitive variables. *Psychology and Aging, 13,* 486–500.
Schaie, K. W. (1988). Internal validity threats in studies of adult cognitive development. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 241–272). New York: Springer-Verlag.
Schaie, K. W., (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.
Schroeder, D. H., & Salthouse, T. A. (2004). Age-related effects in cognition between 20 and 50 years of age. *Personality and Individual Differences, 36,* 393–404.
Sliwinski, M., & Buschke, H. (1999). Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychology and Aging, 14,* 18–33.
Snow, W. G., Tierney, M. C., Zorzitto, M. L., Fisher, R. H., & Reid, D. W. (1989). WAIS-R test–retest reliability in a normal elderly sample. *Journal of Clinical and Experimental Neuropsychology, 11,* 423–428.
Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience, 6,* 389–395.
Thorndike, E. L., Bregman, E. O., Tilton, J. W., & Woodyard, E. (1928). *Adult learning.* New York: Macmillan.
Tisserand, D. J., Visser, P. J., van Boxtel, M. P. J., & Jolles, J. (2000). The relation between global and limbic brain volumes on MRI and cognitive performance in healthy individuals across the age range. *Neurobiology of Aging, 21,* 569–576.
Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist, 3,* 54–56.
Uchiyama, C. L., D'Elia, L. F., Dellinger, A. M., Becker, J. T., Selnes, O. A., Wesch, J. E., et al. (1995). Alternate forms of the Auditory Verbal Learning Test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology, 10,* 133–145.
Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122,* 231–249.
Volkow, N. D., Gur, R. C., Wang, G. J., Fowler, J. S., Moberg, P. J., Ding, Y. S., et al. (1998). Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals. *American Journal of Psychiatry, 155,* 344–349.
Volkow, N. D., Logan, J., Fowler, J. S., Wang, G. J., Gur, R. C., Wong, C., et al. (2000). Association between age-related decline in brain dopamine activity and impairment in frontal and cingulated metabolism. *American Journal of Psychiatry, 157,* 75–80.
WASI Manual (Wechsler Abbreviated Scale of Intelligence). (1999). San Antonio, TX: The Psychological Corporation.
Watson, F. L., Pasteur, M. A. L., Healy, D. T., & Hughes, E. A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology, 9,* 51–61.
Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* New York: The Psychological Corporation.
Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: The Psychological Corporation.
Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., et al. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17,* 179–193.
Yerkes, R. M. E. (1921). Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences, 15,* 1–877.

Received June 9, 2003
Revision received March 11, 2004
Accepted April 29, 2004 ■