Journals of Gerontology: Psychological Sciences
cite as: *J Gerontol B Psychol Sci Soc Sci*, 2016, Vol. 71, No. 1, 49–58
doi:10.1093/geronb/gbu063
Advance Access publication September 2, 2014

Original Research Report

# Aging Cognition Unconfounded by Prior Test Experience

**Timothy A. Salthouse**

Department of Psychology, University of Virginia, Charlottesville.

Correspondence should be addressed to Timothy A. Salthouse, PhD, Department of Psychology, University of Virginia, 102 Gilmer Hall, Charlottesville, VA 22904-4400. E-mail: salthouse@virginia.edu.

Received December 10, 2013; Accepted April 22, 2014

**Decision Editor:** Bob G. Knight, PhD

## Abstract

**Objective.** Investigate time-related age differences in cognitive functioning without influences of prior test experience.
**Methods.** Cognitive scores were compared in different individuals from the same birth years who were tested in different years, when they were at different ages. These types of quasi-longitudinal comparisons were carried out on data from three large projects: the Seattle Longitudinal Study [Schaie, K. W. (2013). Developmental influences on adult intelligence: The Seattle Longitudinal Study (2nd ed.). New York, NY: Oxford University Press], the Betula Project [Ronnlund, M., & Nilsson, L-G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. Intelligence, 36, 192–209], and the Virginia Cognitive Aging Project (this study).
**Results.** In each data set, the results revealed that the estimates of cognitive change with no prior test experience closely resembled the estimates of age relations based on cross-sectional comparisons. Furthermore, longitudinal comparisons revealed positive changes at young ages that gradually became more negative with increased age, whereas all of the estimates of change without prior test experience were negative except those for measures of vocabulary.
**Discussion.** The current results suggest that retest effects can distort the mean age trends in longitudinal comparisons that are not adjusted for experience. Furthermore, the findings can be considered robust because the patterns were similar across three data sets involving different samples of participants and cognitive tests, and across different methods of controlling experience effects in the new data set.

**Key Words:** Aging—Cognition—Longitudinal—Methodology—Retest effects.

Because assessments of cognitive functioning can be reactive and alter the phenomenon under investigation, estimates of cognitive change in longitudinal studies can be distorted by effects associated with the initial test experience. Test experience effects may be responsible for the positive changes often reported at young ages (e.g., Arenberg & Robertson-Tchabo, 1980; Bielak, Anstey, Christensen & Windsor, 2012; Caselli et al., 2009; Giambra, Arenberg, Zonderman, Kawas & Costa, 1995; Huppert & Whittington, 1993, tables 9.7 and 9.11; Ronnlund & Nilsson, 2006, table 3; Rönnlund, Nyberg, Bäckman, & Nilsson, 2005, tables 3 and 4; Salthouse, 2011, 2012, in press; Schaie, 2013, tables 5.1, 5.8, and 5.10; van der Elst, van Boxtel, van Breukelen & Jolles, 2008) and could lead to an underestimation of the magnitude of cognitive decline at all ages if declines are offset by positive test experience effects. Furthermore, correlations of cognitive change may be imprecise if the contributions of experience to change vary across people, but are not taken into consideration in the analyses. Distinguishing the different influences on cognitive change is therefore essential to allow meaningful investigation of the nature of cognitive changes, and their relations with other variables.

One approach that has been used to try to isolate test experience effects from age effects has involved the use of statistical models incorporating different sets of assumptions. For example, one type of model is based on the assumption that if there is variability across participants in the intervals between measurement occasions, the perfect collinearity of the increase in age and the increase in test experience present in longitudinal studies with fixed intervals between occasions can be reduced to allow separate estimates of effects associated with age and effects associated with test experience (e.g., McArdle, Ferrer-Caja, Hamagami & Woodcock, 2002; Salthouse, 2009; Salthouse, Schroeder & Ferrer, 2004). However, unless the range of intervals is very large, the correlation between the increase in age and the increase in test experience will likely still be quite high, and the comparisons involve different people at different intervals who may not be equivalent in all relevant respects.

Another type of statistical model is based on the assumption that when more than two longitudinal assessments are available, different functions can be postulated for the effects of experience and for the effects of age (e.g., Ferrer, Salthouse, McArdle, Stewart & Schwartz, 2005; Ferrer, Salthouse, Stewart, & Schwartz, 2004; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001; Tucker-Drob, Johnson & Jones, 2009; Wilson, Li, Bienias & Bennett, 2006). For example, experience effects could be assumed to asymptote after the second assessment, whereas age effects could be assumed to be linear across all assessments. Although two-function models of this type have the advantage that estimates of experience effects and estimates of age effects can theoretically be derived at the level of individual participants, the assumptions that time-related change is actually composed of two distinct functions, that the functions are of the postulated form (e.g., that experience effects are minimal after the second occasion), and that the functions directly correspond to the effects of age and the effects of test experience, have not yet been independently validated.

These statistical models have often been applied to data from participants across a wide range of ages, but with sufficiently large samples they could also be applied to data from groups with relatively narrow age ranges to provide estimates of age and test experience effects at different ages. However, a limitation of both types of statistical approaches is that the models do not indicate the age trajectory *without* an influence of prior test experience, but instead provide estimates of the trajectory that would be expected if everyone were equated with respect to the amount of prior test experience (cf. Hoffman, Hofer & Sliwinski, 2011).

Another method that has been used to investigate test experience effects in longitudinal studies involves comparing the performance of longitudinal participants on a second occasion with the performance of a comparable sample of people of the same age and tested at the same time on the first occasion, and thus without any prior test experience. This procedure is based on the assumption that the two samples of participants were initially equivalent, sometimes after an adjustment for selective attrition, such that any differences at the time of testing could be attributed to the effects of prior test experience in the longitudinal sample. One of the first descriptions of this method was by Schaie (1988) and it has been used in several studies (e.g., Rönnlund et al., 2005; Ronnlund & Nilsson, 2006), including two with subsets of participants from the current project (i.e., Salthouse, 2009, 2010).

Although the twice-minus-once-tested procedure is valuable for estimating experience effects, it is an indirect method of determining change without test experience because an estimate of the experience effect is first derived, and then it is subtracted from the observed longitudinal change to infer the change that would occur without experience. However, a similar type of reasoning can be used to derive more direct estimates of the change expected without an influence of testing experience. That is, instead of comparing a new sample with the longitudinal sample at the second occasion, new samples of participants from the same birth years could be compared across successive test years when the individuals are at different ages. The contrast of independent samples tested in different years occurs across the same time interval, and hence involves the same increase in age as a longitudinal design, but any differences in performance cannot be affected by prior test experience because all of the participants are only tested once. Furthermore, unlike cross-sectional comparisons, all of the individuals are from the same birth years and thus any age differences cannot be attributed to factors associated with birth cohort (Salthouse, 2013).

Horn and Donaldson (1976) referred to comparisons of people from the same birth year but tested in different years as quasi-longitudinal. Comparisons of this type were reported by Schaie and colleagues (e.g., Schaie, Labouvie & Buech, 1973; Schaie & Strother, 1968), who suggested that the quasi-longitudinal (which they termed independent-samples same-cohort) age trends more closely resembled the age trends in longitudinal comparisons than those in cross-sectional comparisons. However, this interpretation was challenged by Horn and Donaldson (1976) who noted that the age trends for repeated (longitudinal) comparisons were more positive than those for the independent samples (quasi-longitudinal) comparisons. Furthermore, separate estimates of longitudinal and quasi-longitudinal age trends derived from published data of the Seattle Longitudinal Study by Salthouse (1991) suggested that the age trends in the quasi-longitudinal comparisons were nearly identical to the cross-sectional trends, and other studies (e.g., Arenberg, 1978; Kaufman, 2001) have also reported very similar age trends in quasi-longitudinal and cross-sectional comparisons.

Quasi-longitudinal comparisons at different ages can be derived from published data in two large-scale projects in which the same test battery was administered to different samples of adults across a wide range of ages in different years. One project is the Seattle Longitudinal Study (SLS; Schaie, 2013), in which latent variables based on multiple tests of several cognitive abilities were obtained beginning in 1984. Means of the latent variables were reported for independent samples of healthy adults tested in 1984 ($N$ = 629), 1991 ($N$ = 691), and 1998 ($N$ = 719), with the same type of recruitment each year. The relevant data for the computation of quasi-longitudinal trends are reported in table 4.4 of Schaie (2013), which contains means of the latent variables for different cognitive abilities in T-score units. As an example of the quasi-longitudinal contrast, the mean score of 25-year-olds tested in 1984 can be compared with the mean score of 32-year-olds tested in 1991, and with the mean score of 39-year-olds tested in 1998.

The quasi-longitudinal trends in four latent variables from the SLS are portrayed at different ages in Figure 1. Successive points in each line are based on means from different test years, and therefore indicate the quasi-longitudinal change without prior test experience. Data points in the same ordinal position on different lines correspond to cross-sectional comparisons (i.e., the first point in the first line can be compared with the first point in the second line for the cross-sectional difference between 25-year-olds and 32-year-olds in 1984). It can be seen that the quasi-longitudinal age trends were negative at all ages for the measure of verbal memory, and after about 40 for the reasoning and speed measures. However, the age trends were flat in both cross-sectional and quasi-longitudinal comparisons for the verbal comprehension variable.

A second data set that can be used to compute quasi-longitudinal age trends is that from the Betula Project (Ronnlund & Nilsson, 2008), in which independent samples of healthy adults were recruited in 1989, 1994, 1999, and 2004 to perform tests assessing episodic memory, semantic memory, and the Wechsler Block Design test. The same recruitment procedures were used in each test year, with moderately large samples of 100 individuals in each of 10 age groups in 1989 and 1994, and 50 individuals per age group in 1999 and 2004.

Relevant data for the computation of quasi-longitudinal trends are reported as $z$ scores in table 3 in Ronnlund and Nilsson (2008), with each column containing data for successive test years, and each row containing values for a given age group. As an example of the quasi-longitudinal contrast, the first mean in the first column, corresponding to 35-year-olds tested in 1989, can be compared with the second mean in the second column, corresponding to

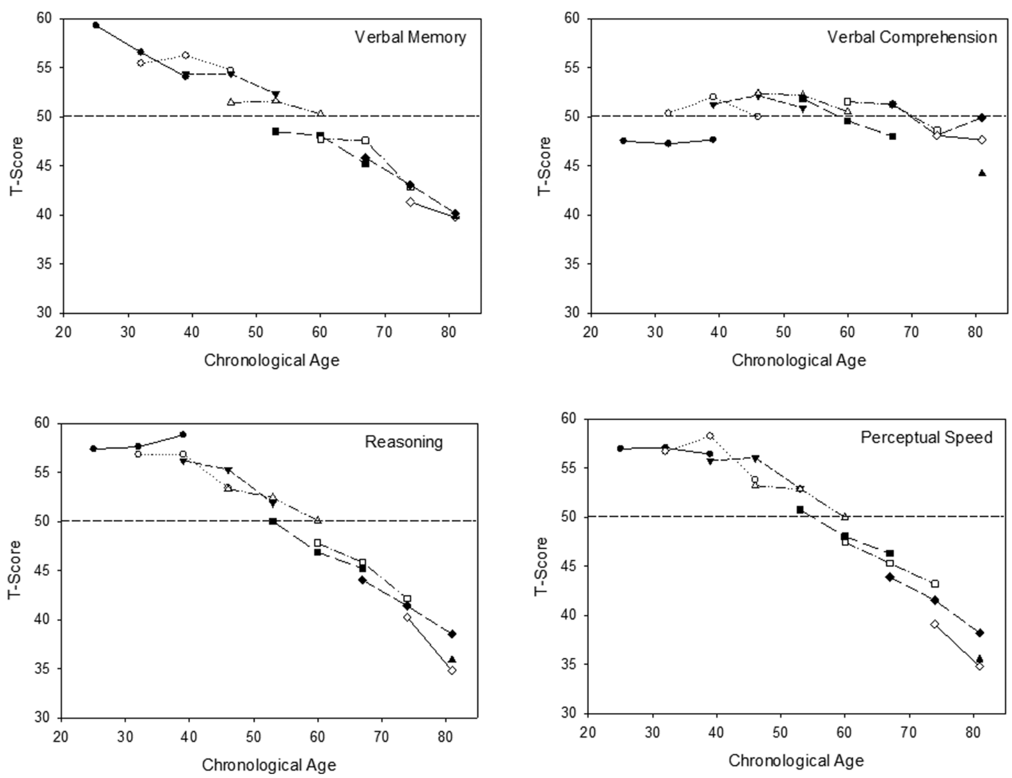


**Figure 1.** Mean scores in three test years for individuals from different birth years for latent variables representing verbal memory, verbal comprehension, reasoning, and perceptual speed in the Seattle Longitudinal Study. Connected lines with the same symbol represent scores from people of the same birth cohort who were tested for the first time in different years when they were at different ages. Data from table 4.4 in Schaie (2013).

40-year-olds tested in 1994, with the third mean in the third column, corresponding to 45-year-olds tested in 1999, and with the fourth mean in the fourth column, corresponding to 50-year-olds tested in 2004. Separate values were reported for men and women, but they were averaged in the current analyses to provide more stable values.

The quasi-longitudinal trends with the episodic memory, semantic memory, and block design scores in the Betula Project are portrayed in the three panels of Figure 2. As in Figure 1, the quasi-longitudinal contrasts are represented by lines connecting the same type of symbol, and the cross-sectional comparisons correspond to points in the same ordinal position on different lines. It can be seen that there were negative time-related differences within each age group with all three cognitive ability measures, and substantial convergence of the age trends in quasi-longitudinal and cross-sectional comparisons.

The patterns in Figures 1 and 2 indicate that the estimates of age-related change in measures of cognitive functioning without an influence of prior test experience closely resemble the cross-sectional age trends. However, a critical assumption of quasi-longitudinal comparisons is that the samples in different test years are equivalent, and neither the SLS nor the Betula Project included direct assessment of the equivalence of the samples in different test years. The primary purpose of the current study was to extend these types of quasi-longitudinal comparisons by using measures of the selectivity of the samples across test years as covariates in the analyses, and by conducting the comparisons with short intervals between test years to minimize any time-related social and cultural differences that might contribute to differential selectivity of participants in different test years. In addition, longitudinal comparisons were included to allow a direct comparison of longitudinal and quasi-longitudinal age trends.

The analyses are based on data from the Virginia Cognitive Aging Project (VCAP; Salthouse, 2007, 2009, in press; Salthouse, Pink & Tucker-Drob, 2008) that is well suited to investigate relations between age and cognitive change because the participants span a wide age range, and the assessment of cognition was broad, with three or four tests representing each of five separate cognitive domains. VCAP started as a cross-sectional study in 2001, and in 2004 it was expanded to include longitudinal assessments, with recruitment of new and returning participants continuing every year since 2004.

It is nearly impossible to obtain a truly random sample of adults, if for no other reason than that people cannot be forced to participate in a research project. Furthermore, quota samples selected to match population proportions in important demographic characteristics are expensive and difficult to obtain. However, because systematic quota sampling is often used to establish representative norms in commercial cognitive test batteries, it is possible to capitalize on those efforts by comparing cognitive scores in a target sample to the scores in the nationally representative normative sample. Four tests from the Wechsler cognitive (WAIS III, Wechsler, 1997a) and memory (WMS III, Wechsler, 1997b)

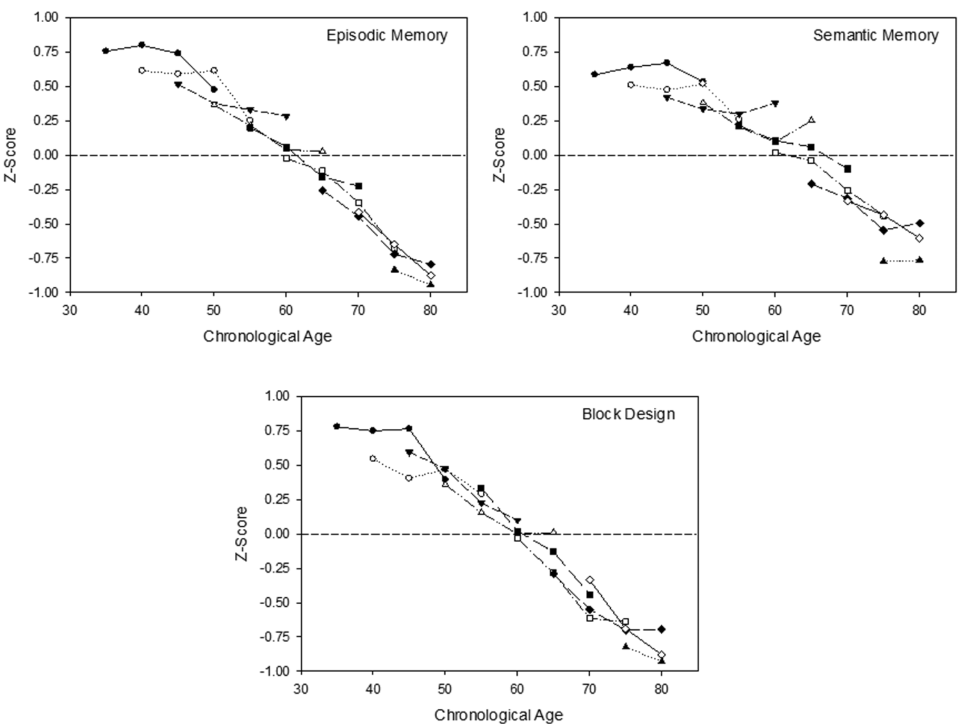


**Figure 2.** Mean scores in four test years for individuals from different birth years for episodic memory, semantic memory, and block design scores in the Betula Project. Connected lines with the same symbol represent scores from people of the same birth cohort who were tested for the first time in different years when they were at different ages. Data from table 3 in Ronnlund and Nilsson (2008).

test batteries were used for this purpose in the current project. In addition, to minimize influences of impending death or late-life health conditions that might affect cognitive functioning and distort age-change relations, only adults under 80 years of age at the first assessment are considered in this report.

To summarize, the present study was designed to investigate age trends in longitudinal and quasi-longitudinal comparisons in different cognitive abilities. It was hypothesized that, consistent with the results in Figures 1 and 2, quasi-longitudinal comparisons would reveal consistently negative age-related differences whereas longitudinal comparisons would reveal positive changes that gradually became more negative with increasing age.

## Method

### Virginia Cognitive Aging Project

Characteristics of the VCAP participants in the quasi-longitudinal and longitudinal comparisons at the first occasion are reported in Table 1. It can be seen that the number of participants per decade ranged from 174 to 969, and that increased age was associated with poorer self-ratings of health, but more years of education. Statistical control of the measures of health and education had very little effect on the pattern of results, and therefore these measures were ignored in subsequent analyses. Supplementary Table 1 contains the means of the age-adjusted scaled scores in the Wechsler tests. The normative samples have scaled score means of 10 and standard deviations of 3, and therefore the VCAP participants can be inferred to have a higher average level of functioning on these tests than the individuals in the nationally representative sample used to establish the norms. However, the standard deviations were close to 3 at all ages, indicating that the level of variability was similar to that in the normative sample. The correlations of the scaled scores with age were positive, particularly in the longitudinal sample, indicating that, relative to their age peers, the older participants were more select in terms of their initial level of cognitive functioning than were the younger participants.

### Cognitive Tests

Cognitive functioning in VCAP was assessed with 16 individually administered cognitive tests designed to represent five cognitive domains (i.e., vocabulary knowledge, inductive reasoning, spatial visualization, verbal memory, and perceptual speed). The tests and their sources are briefly described in the Supplementary Appendix. The tests have been found to have coefficient alpha, test–retest, and alternate forms reliabilities above 7 (Salthouse, 2007, in press; Salthouse et al., 2008; Salthouse & Tucker-Drob, 2008), and to have similar factor structures in different age groups (Salthouse et al., 2008; Soubelet & Salthouse, 2011). Furthermore, the five abilities assessed in VCAP can be inferred to represent major dimensions of cognitive functioning because when considered in the context of these abilities, unique age relations across a broad variety of cognitive tests have been few in number and small in magnitude (Salthouse et al., 2008; Salthouse, Siedlecki & Krueger, 2006). In order to minimize task-specific influences and measurement error, all of the analyses in the current project were conducted on composite scores formed by

**Table 1.** Characteristics of VCAP Samples by Decade

| Decade | 20 | 30 | 40 | 50 | 60 | 70 | Age correlation |
|---|---|---|---|---|---|---|---|
| *N* | | | | | | | |
| Quasi-longitudinal | 728 | 408 | 673 | 969 | 706 | 493 | NA |
| Longitudinal | 227 | 174 | 361 | 544 | 401 | 310 | NA |
| Age | | | | | | | |
| Quasi-longitudinal | 23.2 (3.2) | 34.3 (2.8) | 45.0 (2.9) | 54.4 (3.0) | 64.2 (2.9) | 74.2 (2.9) | NA |
| Longitudinal | 22.4 (3.3) | 34.7 (2.9) | 45.2 (2.9) | 54.5 (2.9) | 64.3 (2.9) | 74.5 (3.1) | NA |
| Proportion female | | | | | | | |
| Quasi-longitudinal | .58 | .71 | .72 | .71 | .66 | .58 | .02 |
| Longitudinal | .60 | .74 | .71 | .72 | .65 | .58 | –.03 |
| Health | | | | | | | |
| Quasi-longitudinal | 2.0 (.9) | 2.1 (.8) | 2.1 (.9) | 2.2 (.9) | 2.1 (.9) | 2.4 (.9) | .11* |
| Longitudinal | 2.0 (.8) | 2.3 (.8) | 2.1 (.9) | 2.1 (.9) | 2.1 (.9) | 2.4 (.9) | .08* |
| Education | | | | | | | |
| Quasi-longitudinal | 14.7 (2.1) | 15.8 (2.8) | 15.4 (2.7) | 15.8 (2.6) | 16.4 (2.8) | 15.9 (2.9) | .18* |
| Longitudinal | 14.1 (1.9) | 15.6 (2.4) | 15.4 (2.5) | 16.0 (2.6) | 16.4 (2.6) | 16.0 (3.0) | .22* |
| T1–T2 interval (years) | | | | | | | |
| Longitudinal | 2.9 (1.7) | 2.9 (1.6) | 3.1 (1.6) | 3.0 (1.5) | 2.9 (1.4) | 2.7 (1.2) | –.05 |

*Notes*. Health is a self-rating on a scale from 1 for excellent to 5 for poor.
**p* < .01.

averaging *z* scores for the test scores representing each of the cognitive abilities.

The same VCAP tests have been administered to similar groups of participants since 2001, which allowed Salthouse (2013) to derive estimates of quasi-longitudinal age trends in six age groups. Birth year and test year were included as simultaneous predictors of cognitive performance in multiple regression analyses, with the age-adjusted scaled scores from four Wechsler tests (Vocabulary, Digit Symbol, Word Recall, and Logical Memory) used as covariates to control for possible differences in sample selectivity across test years. The test-year coefficients in these analyses therefore provide estimates of the effect of one year of age on the relevant measure of cognitive functioning without an influence of prior test experience. Importantly, these estimates are obtained from individuals within a 10-year range of birth years, with birth year used as a simultaneous predictor, and therefore the test-year estimates represent effects of age within the same birth cohort. These estimates were used in the current study, and to provide a contrast of longitudinal and quasi-longitudinal age trends, longitudinal data are also reported based on participants who returned for a second occasion after an interval of approximately 3 years. Although the interval varied across participants, the correlation between interval length and age was only -.05, and thus it was ignored in subsequent analyses.

## Results

The observed longitudinal scores at the first (T1) and second (T2) occasion in each age decade for the composite scores are plotted in Figure 3. The solid lines in the figures represent the observed longitudinal change from the first to the second occasion. As is often found, there was an increase in the scores from the first to the second occasion at younger ages, and a decrease at older ages.

A projected T2 score was derived by multiplying the estimated effect of one year of age (obtained from the quasi-longitudinal comparisons reported in Salthouse, 2013) by the number of years between the T1 and T2 assessments, and then adding this product, which corresponds to estimated time-related change over the T1–T2 interval without prior test experience, to the T1 score of the longitudinal participants. These estimates are plotted as solid triangles in Figure 3 and are connected to the observed T1 scores by dashed lines.

With the exception of vocabulary at the youngest ages, all of the projected T2 values were less positive than the observed T2 values. As was the case in the other data sets portrayed in Figures 1 and 2, the quasi-longitudinal age trends (dashed lines) were very similar to the cross-sectional age trends (successive groups of lines).

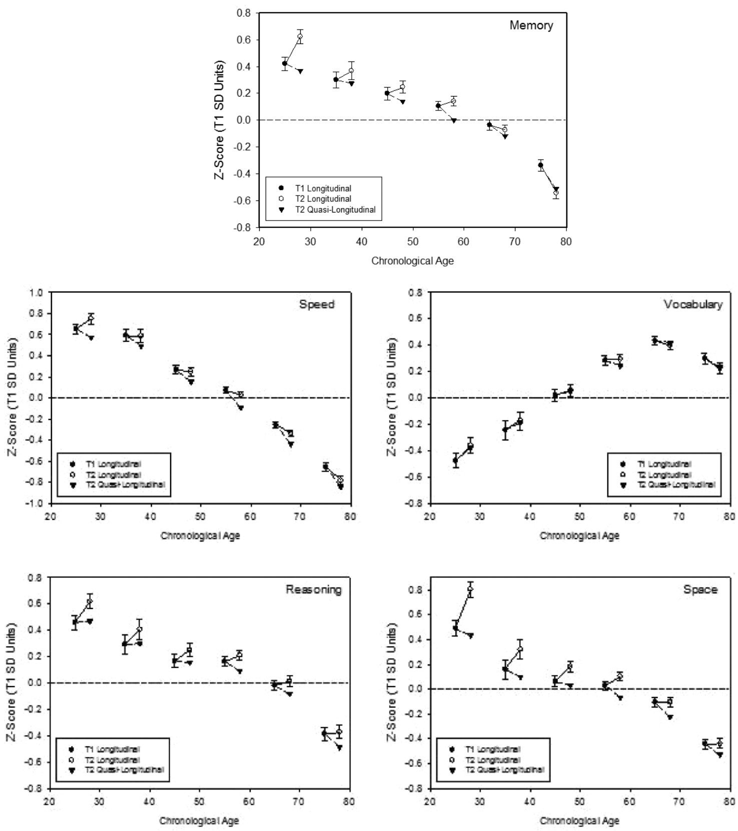


**Figure 3.** Mean T1 and T2 composite scores in the longitudinal sample and projected T2 memory composite score based on the quasi-longitudinal comparison as a function of age decade. The quasi-longitudinal estimates were derived from regression equations relating test score to age in people from the same birth year, who were tested for the first time in different years. Error bars are standard errors.

The difference between the observed (longitudinal) T2 score and the projected (quasi-longitudinal) T2 score can be interpreted as an estimate of the contribution of prior test experience to the change across the T1 to T2 interval. That is, because the projected T2 scores are derived by adding the estimate of experience-independent change to the T1 score, subtraction of the projected T2 score from the observed T2 score provides an estimate of the influence of the initial test experience on the change from T1. These differences are plotted for each ability domain as a function of age decade in Supplementary Figure 1.

One important point to note in Supplementary Figure 1 is that the values were close to zero for vocabulary, which is consistent with the nearly identical longitudinal and quasi-longitudinal trends in Figure 3. It is also noteworthy that the estimates of test experience effects for the speed and reasoning abilities were all positive and nearly constant at all ages, whereas with the memory and space abilities the estimates of test experience effects were much more positive at younger ages than at older ages.

A final analysis used methods similar to those described in Salthouse (2010) to predict the score at T2 based on the twice-minus-once-tested comparison. Specifically, a predicted T2 value was derived from scores of individuals of the same age and tested at the same time as the longitudinal participants on the second occasion, after adding the difference between the T1 scores of the longitudinal sample and the T1 scores of entire sample to adjust for selective attrition. These estimates for the memory composite scores are portrayed in Supplementary Figure 2 along with quasi-longitudinal estimates and the observed T1 and T2 scores from the longitudinal samples. The cross-sectional age trends in these data are represented by dotted lines connecting the T1 values.

Despite the differences in their derivation (i.e., the differences across independent samples tested in different years are added to the value of the longitudinal sample at T1 for the quasi-longitudinal method, and the value of the new sample tested at the same age and time as the T2 longitudinal sample after an adjustment for initial selectivity in the twice-minus-once method), it is noteworthy that most of the predicted T2 values from the twice-minus-once method are close to the values from the quasi-longitudinal method. Furthermore, for adults under about 60 years of age, both of the predicted T2 scores after adjusting for prior test experience were more closely aligned with the cross-sectional age relations represented by the dotted lines than with the observed longitudinal T2 scores. Although not portrayed to save space, very similar patterns were evident in analyses of the other cognitive abilities.

## Discussion

As is apparent in Figure 3, longitudinal comparisons often reveal an increase in scores on cognitive tests from the first to the second occasion for adults under about 60 years of age. One interpretation of the positive change at young ages is that it reflects benefits of prior experience with the tests. However, little is known about the magnitude of these test experience effects, and whether they operate only in young adults, or across all of adulthood. Quasi-longitudinal comparisons are relevant to these issues because they provide estimates of the magnitude of age-related cognitive changes that occur without any influences of prior test experience (Schaie et al., 1973). The comparisons are based on the assumptions that time-related differences in both longitudinal and quasi-longitudinal comparisons reflect influences of age, of period effects associated with physical and cultural conditions at the time of assessment, and of possible shifts associated with the nature of the assessment (e.g., identity of the experimenters, location of the test site, individual vs. group examination, paper-and-pencil vs. computer administration). The primary differences in the two types of designs are that the same individuals contribute data at each occasion in longitudinal studies, and thus the comparisons can be influenced by prior test experience, whereas all of the measurements in quasi-longitudinal comparisons are obtained from different individuals, and hence the contrasts cannot be affected by prior test experience. If these assumptions are valid, and if the samples in each test year are comparable, quasi-longitudinal comparisons provide an estimate of the age-related cognitive change that occurs without an influence of prior test experience.

Similar types of comparisons were carried out in three different data sets, which differed in the nature of the cognitive tests, the intervals between test years (i.e., 1 year in the current project, 5 years in the Betula Project, and 7 years in the SLS), the number of test years (i.e., 10 in the current project, 4 in the Betula Project, and 3 in the SLS), and the time period of the testing (i.e., from 2001 to 2011 in the current project, from 1989 to 2004 in the Betula Project, and from 1984 to 1998 in the SLS). Despite numerous procedural differences, the overall patterns were similar in all three data sets. In each case the estimates of experience-independent change were either stable or negative at all ages, and closely approximated the estimated age trends based on cross-sectional comparisons.

There were some differences across data sets, such as with the vocabulary construct in which there was positive change in VCAP, stability in SLS, and negative change for the similar semantic memory construct in the Betula Project. At least some of these differences may be attributable to the use of different tests to represent the construct, as a speeded verbal fluency test was included in the semantic memory construct in the Betula Project, whereas only knowledge tests were used in the other projects.

In order to provide more direct comparisons across ability domains and data sets, regression equations were computed relating the ability measure (in *z*-score units) to test year (and age) in narrow age groups. If the time-related effects are assumed to be primarily linear, the test-year slopes of these equations provide an estimate of the amount of (experience-independent) cognitive change per year of age. The SLS data were reported in T-score units, and therefore they were converted to *z* scores by subtracting 50 (the T-score mean) and dividing by 10 (the T-score standard deviation). The values of the estimated change per year without an influence of prior test experience are reported in Supplementary Table 2, and the estimates of change in episodic memory from the three data sets are plotted in Supplementary Figure 3. Inspection of the table and the figure indicates that, at least between 30 and 70 years of age, the estimates in the three data sets were quite consistent with one another. For example, all of the estimates of annual memory decline between 65 and 67 years of age were between .03 and .04 standard deviation units per year. The similar estimates are particularly noteworthy because different tests were used to represent the memory construct in the three data sets. That is, tests of word recall, story memory, and paired associates memory were used in VCAP; tests of immediate and delayed memory of a 20-word list were used in SLS; and tests of recall and recognition of statements of actions and recall of fictitious facts were used in the Betula Project.

In addition to the replication of the patterns in different samples with the same analytical procedure, as portrayed in Figures 1 and 2, and Supplementary Figure 3, the results in Supplementary Figure 2 indicate that estimates of experience-independent change were similar to the quasi-longitudinal and twice-minus-once-tested procedures. Although both methods are based on comparisons of samples of participants tested for the first time, they differ in the specific groups that are compared and in the relevant contrasts. That is, the quasi-longitudinal estimates are based on the performance differences across samples from the same birth cohorts tested in different years, whereas the twice-minus-once-tested estimates are based on a sample tested for the first time when the longitudinal sample was tested for the second time. The convergence of the results across different samples and different methods increases confidence in the conclusion that the direction and magnitude of cognitive change is distorted in conventional longitudinal studies, especially at younger ages.

It is important to note that these results are inconsistent with the conclusions of Schaie and colleagues (1973) that "there are dramatic discrepancies between cross-sectional and within-cohort age differences (p. 162)," and that "independent measurements to assess within-cohort change patterns are in almost perfect agreement with the repeated measurement data (p. 164)." However, the current results are consistent with the patterns in Figure 1 through 10 of Schaie and colleagues (1973), and with the results of analyses of more recent data from that project portrayed in Figure 1.

Estimates of the magnitude of the test experience effects can be derived from the comparisons of the longitudinal and quasi-longitudinal differences plotted in Supplementary Figure 1. The fact that most of the values were positive implies that estimates of change from longitudinal comparisons underestimate the negative change that would have occurred without prior test experience. Furthermore, the estimated test experience effects were constant across the entire age range for speed and reasoning, which suggests that the observed longitudinal change may be more positive than the experience-independent change at every age, without altering the relations between age and change. However, with memory, and to a lesser extent with spatial ability, the age trends in longitudinal comparisons may be distorted relative to what might be expected without prior test experience because the experience effects were larger at young ages than at older ages.

One limitation of the current research is that the estimates were derived from a particular set of cognitive tests, and an average retest interval of about 3 years. Although different test experience effects might be apparent with other tests or at different intervals, this concern is alleviated somewhat by the similar pattern in the other two data sets involving different combinations of tests and retest intervals. Another limitation of the study is that it is not known whether test experience effects such as those portrayed in Supplementary Figure 1 only affect change from a first to a second assessment, or whether they would also be evident on later assessments. In addition, because the analyses are based on differences at the level of group means, little is known about possible individual differences in the magnitude of test experience effects.

In summary, although influences of test experience on longitudinal change are widely acknowledged, the current results are novel in indicating that negative change is apparent at relatively young ages in several cognitive domains. In three separate and moderately large data sets, the age trends based on quasi-longitudinal comparisons closely resemble the age trends based on cross-sectional comparisons. Although the results of the quasi-longitudinal comparisons imply that the age trends in longitudinal comparisons reflect test experience effects in addition to aging effects, and therefore may underestimate the magnitude of cognitive decline, it is important to emphasize that only longitudinal comparisons are useful for examining individual differences in change, and for investigating correlates of individual differences in change. That is, there is no substitute for longitudinal data if one is interested in phenomena occurring within, as opposed to between, individuals. Nevertheless, because measures of within-person change incorporate effects of prior test experience in addition to effects of aging, a high priority for future research should be the development of valid methods of distinguishing experience and age effects at the level of individual participants, possibly by combining results from longitudinal comparisons with results from the types of quasi-longitudinal comparisons reported here.

## Supplementary Material

Supplementary material can be found at: http://psychsocgerontology.oxfordjournals.org/

## Funding

The project was supported by award number R37AG024270 from the National Institute on Aging. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

## References

Arenberg, D. (1978). Differences and changes with age in the Benton Visual Retention Test. *Journal of Gerontology*, **33**, 534–540. doi:10.1093/geronj/33.4.534

Arenberg, D. & Robertson-Tchabo, E. A. (1980). Age differences and age changes in cognitive performance; New "old" perspectives. In R. L. Sprott (Ed.), *Age, learning, ability and intelligence* (pp. 139–157). New York: Van Nostrand Reinhold.

Bielak, A. A. M., Anstey, K. J., Christensen, H., & Windsor, T.D. (2012). Activity engagement is related to level, but not change in cognitive ability across adulthood. *Psychology and Aging*, **27**, 219–228. doi:10.1037/a0024667

Caselli, R. J., Dueck, A. C., Osborne, D., Sabbagh, M. N., Connor, D. J., Ahern, G. L., …Reiman, E. M. (2009). Longitudinal modeling of age-related memory decline and the APOE e4 effect. *The New England Journal of Medicine*, **361**, 255–263.

Ferrer, E., Salthouse, T. A., McArdle, J. J., Stewart, W. F., & Schwartz, B. S. (2005). Multivariate modeling of age and retest in longitudinal studies of cognitive abilities. *Psychology and Aging*, **20**, 412–422. doi:10.1037/0882-7974.20.3.412

Ferrer, E., Salthouse, T. A., Stewart, W., & Schwartz, B. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging*, **19**, 243–259. doi:10.1037/0882-7974.19.2.243

Giambra, L. M., Arenberg, D., Zonderman, A. B., Kawas, C. & Costa, P. T. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychology and Aging*, **10**, 123–139. doi:10.1037/0882-7974.10.1.123

Hoffman, L., Hofer, S. M. & Sliwinski, M. J. (2011). On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: A simulation study. *Psychology and Aging*, **26**, 778–791. doi:10.1037/a0023910

Horn, J. L. & Donaldson, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist*, **31**, 701–719. doi:10.1037/0003-066X.31.10.701

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). Aldershot, England: Dartmouth.

Kaufman, A. S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence*, **29**, 131–167. doi:10.1016/S0160-2896(00)00046-5

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. doi:10.1037/0012-1649.38.1.115

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. doi:10.1016/S0028-3932(00)00099-3

Ronnlund, M., & Nilsson, L-G. (2006). Adult life-span patterns in WAIS–R Block Design performance: Cross-sectional versus longitudinal age gradients and relations to demographic factors. *Intelligence*, **34**, 63–78. doi:10.1016/j.intell.2005.06.004

Ronnlund, M., & Nilsson, L-G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. *Intelligence*, **36**, 192–209. doi:10.1016/j.intell.2007.05.002

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, **20**, 3–18. doi:10.1037/0882-7974.20.1.3

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Salthouse, T. A. (2007). Implications of within-person variability in cognitive and neuropsychological functioning on the interpretation of change. *Neuropsychology*, **21**, 401–411. doi:10.1037/0894-4105.21.4.401

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, **30**, 507–514. doi:10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, **24**, 563–572. doi:10.1037/a0019026

Salthouse, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin*, **137**, 753–784. doi:10.1037/a0023262

Salthouse, T. A. (2012). Does the direction and magnitude of cognitive change depend on initial level of ability? *Intelligence*, **40**, 352–361. doi:10.1016/j.intell.2012.02.006

Salthouse, T. A. (2013). Within-cohort age differences in cognitive functioning. *Psychological Science*, **24**, 123–130. doi:10.1177/0956797612450893

Salthouse, T. A. (in press). Correlates of cognitive change. *Journal of Experimental Psychology: General*.

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, **36**, 464–486. doi:10.1016/j.intell.2007.10.003

Salthouse, T. A., Schroeder, D. H., & Ferrer, E. (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Developmental Psychology*, **40**, 813–822. doi:10.1037/0012-1649.40.5.813

Salthouse, T. A., Siedlecki, K. L. & Krueger, L. E. (2006). An individual differences analysis of memory control. *Journal of Memory and Language*, **55**, 102–125. doi:10.1016/j.jml.2006.03.006

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*, (pp. 241–272). NY: Springer-Verlag. doi:10.1007/978-1-4612-3852-2_8

Schaie, K. W. (2013). *Developmental influences on adult intelligence: The Seattle Longitudinal Study* (2nd ed.). New York, NY: Oxford University Press.

Schaie, K. W., Labouvie, G. V., & Buech, B. U. (1973). Generational and cohort-specific differences in adult cognitive functioning: A fourteen-year study of independent samples. *Developmental Psychology*, **9**, 151–166. doi:10.1037/h0035093

Schaie, K. W., & Strother, C. R. (1968). The effect of time and cohort differences on the interpretation of age changes in cognitive behavior. *Multivariate Behavioral Research*, **3**, 259–293. doi:10.1207/s15327906mbr0303_1

Soubelet, A. & Salthouse, T. A. (2011). Correlates of level and change in the Mini-Mental Status Exam. *Psychological Assessment*, **23**, 811–818. doi:10.1037/a0023401

Tucker-Drob, E. M., Johnson, K. E. & Jones, R. N. (2009). The cognitive reserve hypothesis: A longitudinal examination of age-associated declines in reasoning and processing speed. *Developmental Psychology*, **45**, 431–446. doi:10.1037/a0014012

Van der Elst, W., van Boxtel, M. P. J., van Breukelen, G. J. P., & Jolles, J. (2008). Detecting the significance of changes in performance on the Stroop Color-Word Test, Rey's Verbal Learning Test, and the Letter-Digit Substitution Test: The regression-based change approach. *Journal of the International Neuropsychological Society*, **14**, 71–80. doi:10.1017/S1355617708080028

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.

Wilson, R. S., Li, Y., Bienias, J. L. & Bennett, D. A. (2006). Cognitive decline in old age: Separating retest effects from the effects of growing older. *Psychology and Aging*, **21**, 774–789. doi:10.1037/0882-7974.21.4.774