Why are there different age relations in cross-sectional and longitudinal comparisons of cognitive functioning?

Timothy A. Salthouse

Department of Psychology, University of Virginia, Charlottesville, VA 22904-4400

Timothy A. Salthouse: Salthouse@virginia.edu

Abstract

A major challenge for researchers interested in investigating relations between aging and cognitive functioning is distinguishing influences of aging from other determinants of cognitive performance. For example, cross-sectional comparisons may be distorted because people of different ages were born and grew up in different time periods, and longitudinal comparisons may be distorted because performance on a second occasion is influenced by the experience of performing the tests on the first occasion. One way in which these different types of influences might be investigated is with research designs involving comparisons of people of different ages from the same birth cohorts who are all tested for the first time in different years. Results from several recent studies using these types of designs suggest that the age trends in some cognitive abilities more closely resemble those from cross-sectional comparisons than those from longitudinal comparisons. These findings imply that a major reason for different age trends in longitudinal and cross-sectional comparisons of cognitive functioning is that the prior experience with the tests inflates scores on the second occasion in longitudinal studies.

Cross-sectional comparisons involving different people at each age typically reveal nearly monotonic declines in several measures of cognitive functioning beginning when people are in their 20's (Salthouse, 2009; Schaie 2013). However, age trends in longitudinal comparisons involving the same people at different ages often reveal maintained, or even increased, levels of performance (e.g., Ronnlund & Nilsson, 2006; Ronnlund et al. 2005; Salthouse, 2009; Schaie, 2013).

Two major hypotheses have been proposed to explain the discrepant age trends in cross-sectional and longitudinal comparisons. One hypothesis is that cross-sectional comparisons are misleading because people of different ages may differ in factors other than age that could be contributing to different levels of cognitive performance. Because it has been difficult to identify all of the relevant characteristics that could differ across people of different ages, birth year (or birth cohort) is often used as a proxy for generation-specific influences on cognitive functioning. This interpretation is frequently expressed in the form of a proposal that cross-sectional age differences are confounded by cohort differences (e.g., Schaie, 2009).

A second hypothesis that has been proposed to account for the discrepant age trends in cross-sectional and longitudinal designs is that longitudinal comparisons are distorted because performance on a second occasion can be influenced by prior experience with the

tests. That is, because cognitive testing can be reactive, such that performance on a subsequent assessment is often higher than it would have been without an initial assessment, this interpretation postulates that longitudinal comparisons may yield misleading age trends because they are affected by practice, or retest, effects.

Although the relation between age and cognition is one of the most fundamental issues in the field of cognitive aging, surprisingly little research has directly investigated the relative contributions of cohort and experience on age trends in measures of cognitive functioning. However, two methods originally proposed by Warner Schaie have been used in a number of studies to investigate experience effects in individuals from the same birth cohorts. The two methods are schematically illustrated in Figure 1 in the context of predicting the cognitive score at the second (T2) occasion. Notice that a minimum of three samples of research participants are involved in each type of analysis, consisting of the longitudinal sample at two occasions (S1a and S1b), and two separate samples from the same birth years as the longitudinal sample who are tested in the same years as the first (S2) and second (S3) longitudinal occasions. In order to be concrete, specific test years and ages are indicated in the example, but the methods are applicable across a wide range of test years and ages.

The twice-minus-once-tested method is based on performance of a new sample of participants at T2 (S3) serving as an estimate of the score on the second occasion that would have been expected in the longitudinal participants if they had no prior test experience. However, because people who return for a second occasion often have higher scores at the first occasion than those who do not return, the S3 score is adjusted for selectivity of the longitudinal sample (S1a) relative to the total sample (i.e., S1a + S2) at the first occasion. That is, the difference between the score of the longitudinal sample (S1a) and the score of the total sample (S1a and S2) at the initial occasion is added to the score of the S3 sample to account for the selectivity of returning participants.

The quasi-longitudinal method is based on the difference in scores between two new samples of participants (S2 and S3) who are tested in different years, when they were at different ages. This difference, which can be considered an estimate of the change expected without prior test experience, is subtracted from the T1 score of the longitudinal sample (T1a) to link the predicted T2 score to the longitudinal sample.

Note that the two methods rely on the same logic of estimating experience-independent change from the difference between separate samples of adults who are tested in different years. Unlike cross-sectional comparisons, both methods involve comparisons of people from the same birth cohort, and thus the comparisons are not distorted by generation-specific effects, but unlike longitudinal comparisons, the methods provide estimates of change without prior test experience because the comparisons are based on each participant's initial assessment. If the samples in different years are assumed to be comparable in relevant respects, both methods can be viewed as providing estimates of the age trends that would have resulted if there were no confounds of either cohort differences or test experience effects.

The twice-minus-once-tested procedure has been used in several studies (e.g., Schaie, 1988; Ronnlund et al., 2005; 2006; Salthouse, 2009, 2010). The estimates of change without experience varied across studies, possibly because they involved different measures of cognitive functioning and different intervals between assessments (i.e., 2 to 3 years in the Salthouse reports, 5 years in the Ronnlund studies, and 7 years in the Schaie study). However, in nearly every case the estimates of cognitive change without prior experience were less positive than the observed longitudinal changes.

Quasi-longitudinal comparisons were apparently first reported by Schaie (Schaie et al., 1973; Schaie & Strother, 1968), who suggested that the results were similar to those from longitudinal comparisons. However, Schaie's interpretations were challenged by Horn and Donaldson (1981) and by Salthouse (1991), who argued that the quasi-longitudinal results more closely resembled the trends in cross-sectional comparisons than those in longitudinal comparisons. Additional studies using a similar rationale were reported by Arenberg (1978), and Kaufman (2001, 2013). The analyses by Kaufman (2013) were particularly interesting because they capitalized on the samples used to establish the norms for different versions of the Wechsler Adult Intelligence Scale (WAIS) test batteries to estimate age-related cognitive change without prior experience. To illustrate, the WAIS III test battery was normed in 1995 and the WAIS IV test battery was normed in 2007, and therefore individuals born between 1941 and 1950 had a median age of 49.5 in 1995, and a median age of 61.5 in 2007. Comparing the index scores of individuals in these groups (as in Table 7.9 in Kaufman, 2013) thus provides an estimate of the age-related change that might be expected across the 12-year interval without prior test experience. The major result of the Kaufman analyses was that the age gradients in the quasi-longitudinal comparisons closely mirrored the gradients from traditional cross-sectional comparisons.

Application of the two methods in the same data set can be illustrated with published results on a composite memory measure from the Betula project (Ronnlund et al., 2005). This project is particularly relevant in the current context because cross-sectional assessments with independent samples were collected in 1989 and 1994, and a longitudinal follow-up of the original sample was carried out in 1994. Episodic memory was assessed in the Betula project with locally developed tests of recall and recognition of information from sentences describing actions, some of which were performed by the participant. An episodic memory factor score was derived from the individual tests, and the scores in all samples were converted into T-score units based on the means and standard deviation at the first occasion (T1).

Table 1 summarizes relevant results for adults born in 1949 who were tested in 1989 when they were 40 years of age, and tested in 1994 when they were 45 years of age. Note that the cross-sectional comparisons revealed a moderately large age difference favoring the younger group (i.e., age differences of –0.99 in 1989, and –2.40 in 1994), but a positive difference (i.e., +2.08) in the longitudinal comparison. The predicted T2 scores from the twice-minus-once-tested and quasi-longitudinal methods are presented in the bottom of Table 1, along with the estimates of change without experience obtained by subtracting the predicted T2 value from the observed T1 value. The analyses in this example are crude because they are based on only two measurement occasions, the total sample at T1 was used to approximate

the S2 sample, and group means were used rather than scores of individual participants. Under these simplifying conditions, the estimates of change without experience were identical (i.e., −1.26) for the twice-minus-once-tested and quasi-longitudinal methods, and both were more negative than the longitudinal changes (i.e., + 2.08).

Results with the two methods can also be examined with data from the Virginia Cognitive Aging Project (VCAP), which is an on-going longitudinal study of cognitive functioning involving multiple tests of multiple cognitive abilities (Salthouse, 2009, 2011; under review). As of December 2013, over 2,300 individuals have participated in at least two longitudinal occasions, with an average interval between occasions of approximately 3 years. A unique feature of VCAP is that new samples of participants spanning a wide age range have been recruited every year, which allows quasi-longitudinal estimates of change without experience to be derived from regression equations relating age to cognitive performance for individuals within the same birth years (Salthouse, 2013). The recruiting methods were similar in each test year, and scaled scores on the Wechsler tests were also used as covariates in the regression equations to enhance the comparability of samples in different test years.

Analyses in the current article were based on composite scores formed from the average of z-scores (based on the distribution at T1) from three different types of tests (i.e., recall of a list of unrelated words, recall of details of a story, and remembering arbitrary associations between unrelated words). Means of the observed longitudinal scores at T1 and T2 and the predicted score at T2 from the two methods of estimating change without experience are reported in Figure 2.

Three major results should be noted about this figure. First, as is typically found, the cross-sectional age trends (dashed lines) were negative beginning from the youngest age. Second, the longitudinal changes were positive at young ages and gradually became more negative with increased age. And third, age trends of the estimates of change without prior test experience from individuals within the same birth cohorts were more similar to the cross-sectional age trends than to the longitudinal age trends.

Because participants in VCAP perform tests assessing five distinct cognitive domains, results of the twice-minus-once-tested and quasi-longitudinal methods were examined in different cognitive domains (Salthouse, under review). Results with measures of reasoning, spatial visualization, and perceptual speed resembled those in Figure 2, but the age trends with vocabulary were positive until participants were in their 70s, and there was little effect of prior test experience on vocabulary scores at any age.

To summarize, results across several studies indicate that longitudinal comparisons of cognitive functioning are influenced by test experience effects, and that when those effects are eliminated, the age trends closely resemble cross-sectional age trends. Because the relevant samples in the twice-minus-once-tested and quasi-longitudinal samples are from the same birth years, the age trends with these methods cannot be attributed to birth cohort factors.

Although these results are more consistent with the hypothesis that longitudinal comparisons are distorted by experience effects than with the hypothesis that cross-sectional comparisons are distorted by cohort effects, those are not the only interpretations of the different age trends in cross-sectional and longitudinal comparisons. For example, because the people at each age in cross-sectional comparisons are different, observed differences in cognitive functioning could be attributable to characteristics of the individuals other than age, such as quality or quantity of education, exposure to different cultural experiences, etc. This possibility should continue to be explored, but it is important that potentially relevant characteristics are investigated directly instead of relying on birth year as a proxy for possible differences in people of different ages.

Because socio-cultural changes can affect performance on cognitive tests (cf. Flynn effect), age comparisons in cognitive functioning could also be influenced by period effects. Based on time-related improvements in cognitive performance, it is sometimes suggested that cross-sectional comparisons are not meaningful because older adults at the current time may not have been equivalent to the current sample of young adults when the older adults were at that age. However, the critical issue with respect to the interpretation of age differences in cognition is not whether there have been time-related effects on cognitive performance, but whether those effects varied as a function of age. That is, if the period effects were similar at all ages, then relative age comparisons would still be meaningful despite time-related shifts in the absolute level of performance. In contrast, the validity of most longitudinal comparisons could be compromised by time-lag effects because some of the observed change in the score from T1 to T2 may be attributable to period effects, and not to maturational changes of primary interest. The observed longitudinal contrasts might be adjusted for period effects if time-lag data are available at different ages, but there have been few attempts of this type (cf. Salthouse, 1991).

Although the results reported here suggest that the mean age trends in longitudinal comparisons with certain cognitive abilities are distorted by prior test experience, the findings should not be interpreted as diminishing the importance of longitudinal studies. In fact, longitudinal data are absolutely essential for examining change within the same individual, and for identifying relations of other variables with change. However, longitudinal changes may be more accurately estimated, and potentially yield stronger correlations with various lifestyle or neurobiological variables if influences of test experience on cognitive change are removed with methods such as those described here.

In conclusion, several factors are likely contributing to different age trends in cross-sectional and longitudinal comparisons. The groups in cross-sectional comparisons could differ in aspects related to cognitive functioning, and time-related effects associated with the period of measurement could be influence longitudinal comparisons. Nevertheless, the results summarized in this report indicate that longitudinal comparisons can be affected by test experience, and the similar age trends in between-cohort and within-cohort comparisons suggest that birth cohort is not a critical factor contributing to the differences in at least some cognitive abilities.

References

Arenberg D. Differences and changes with age in the Benton Visual Retention Test. Journal of Gerontology. 1978; 33:534–540. [PubMed: 752043]

Horn JL, Donaldson G. On the myth of intellectual decline in adulthood. American Psychologist. 1976; 31:701–719. [PubMed: 1008331]

Kaufman AS. WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. Intelligence. 2001; 29:131–167.

Kaufman, A. Clinical applications II: Age and intelligence across the adult life span. In: Lichtenberg; Kaufman, editors. Essentials of WAIS IV Assessment. 2013.

Ronnlund M, Nilsson LG. Adult life-span patterns in WAIS–R Block Design performance: Cross-sectional versus longitudinal age gradients and relations to demographic factors. Intelligence. 2006; 34:63–78.

Ronnlund M, Nyberg L, Backman L, Nilsson LG. 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. 2005; 20:3–18. [PubMed: 15769210]

Salthouse, T. Theoretical perspectives on cognitive aging. Hillsdale, NJ: Erlbaum; 1991.

Salthouse TA. When does age-related cognitive decline begin? Neurobiology of Aging. 2009; 30:507–514. [PubMed: 19231028]

Salthouse TA. Influence of age on practice effects in longitudinal neurocognitive change. Neuropsychology. 2010; 24:563–572. [PubMed: 20804244]

Salthouse TA. Within-cohort age differences in cognitive functioning. Psychological Science. 2013; 24:123–130. [PubMed: 23319401]

Salthouse, TA. Aging cognition unconfounded by prior test experience. (under review)

Schaie, KW. Developmental influences on adult intelligence: The Seattle Longitudinal Study. 2. New York, NY: Oxford University Press; 2013.

Schaie, KW. Internal validity threats in studies of adult cognitive development. In: Howe, ML.; Brainerd, CJ., editors. Cognitive development in adulthood: Progress in cognitive development research. New York: Springer-Verlag; 1988. p. 241-272.

Schaie KW, Labouvie GV, Buech BU. Generational and cohort-specific differences in adult cognitive functioning: A fourteen-year study of independent samples. Developmental Psychology. 1973; 9:151–166.

Schaie KW, Strother CR. A cross-sequential study of age changes in cognitive behavior. Psychological Bulletin. 1968; 70:671–680. [PubMed: 5712311]

Recommended Readings

Salthouse, T. Theoretical perspectives on cognitive aging. Hillsdale, NJ: Erlbaum; 1991. Chapter 3 in this book contains an early review of the different age trends in cross-sectional and longitudinal designs

Salthouse, TA. Major Issues in Cognitive Aging. NY: Oxford University Press; 2010. Chapter 2 in this book discusses cohort and period influences in more detail than in the current paper

Schaie, KW. Developmental influences on adult intelligence: The Seattle Longitudinal Study. 2. New York, NY: Oxford University Press; 2013. The most recent summary of the classic Seattle Longitudinal Study, with Chapters 6 and 8 containing descriptions of Schaie's procedures to investigate change without prior experience
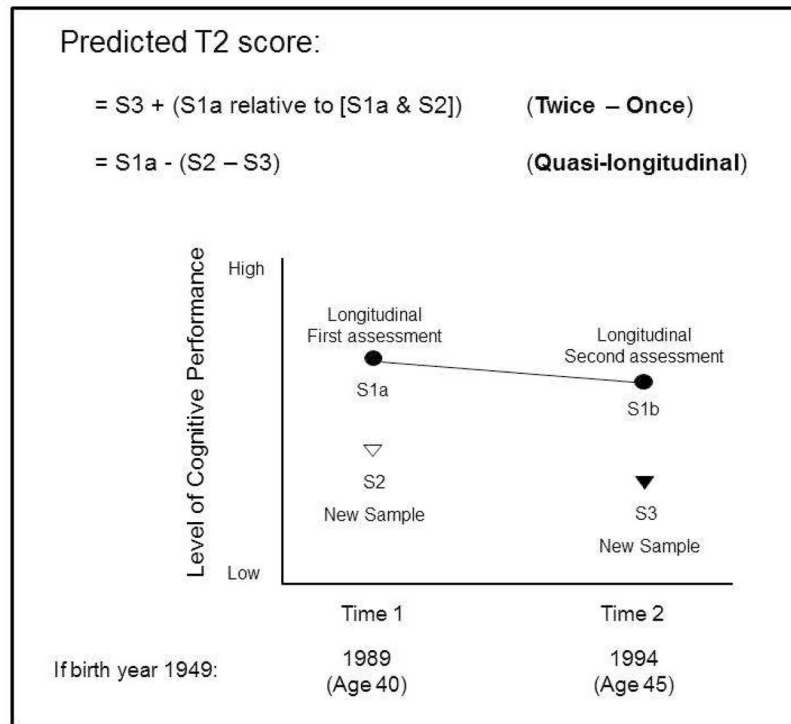
Figure 1.
Schematic illustration of methods that can be used to estimate age-related change without influences of prior test experience. See text for details.
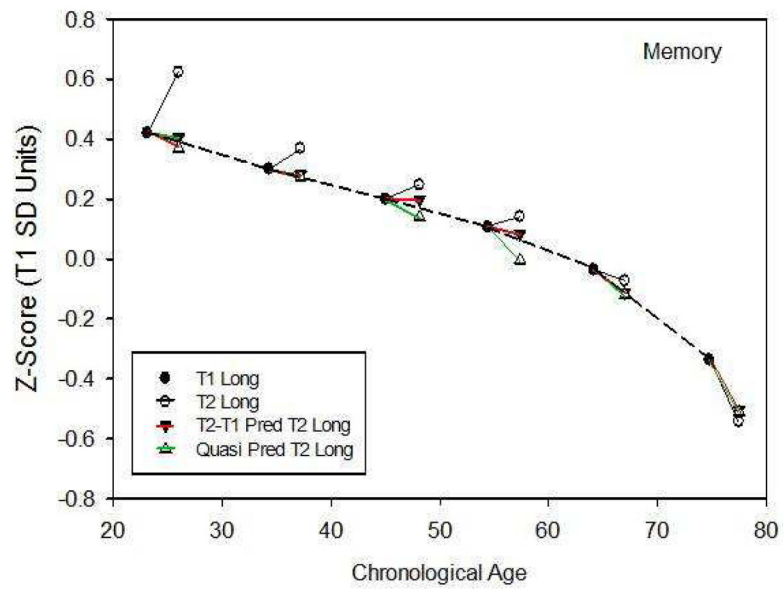
Figure 2.
Mean composite memory scores on the first (T1) and second (T2) longitudinal occasion and predicted second occasion scores based on the twice-minus-once-tested and quasi-longitudinal procedures. The dotted lines are the cross-sectional age trends based on between-cohort comparisons, and the three sets of solid lines are within-cohort comparisons.

Table 1

Illustration of methods to estimate change without experience with data from Ronnlund et al. (2005), Tables 3 and 5

	Cross-Sectional		Longitudinal	
	1989	**1994**	**1989**	**1994**
Age 40	56.40 *(S2)*	57.54	56.41 *(S1a)*	
Age 45	55.41	55.14 *(S3)*		58.49 *(S1b)*
Difference	−0.99	−2.40		+2.08

Predicted T2	**Est. Change from T1 Long (56.41)**
Twice – Once Tested = S3 + (S1a − Avg[S1a+S2])	
55.14 + (56.41 − Avg[56.41+ 56.40]) = 55.15	−1.26
Quasi-Longitudinal = S1a − (S2 − S3)	
56.41 − (56.40 − 55.14) = 55.15	−1.26