Published in final edited form as:
Neuropsychology. 2010 March ; 24(2): 273–278. doi:10.1037/a0017284.

Does the Meaning of Neurocognitive Change Change With Age?

Timothy A. Salthouse
Department of Psychology, University of Virginia, Charlottesville, Virginia

Abstract

Significant declines in longitudinal comparisons of neurocognitive performance are seldom evident until adults are in their 60s or older, but relatively little is known about the existence, or nature, of age-related changes at earlier periods in adulthood. The current research was designed to address this issue by examining characteristics of change in measures from 12 neuropsychological and cognitive tests at different periods in adulthood. Although change was largely positive for adults under about 55 years of age and frequently negative for adults at older ages, the reliabilities of the changes in the neuropsychological and cognitive variables were similar at all ages. Furthermore, there were few systematic relations of age on the reliability-adjusted correlations between the changes in composite scores representing different abilities. These results imply that although neurocognitive declines may not be apparent at young ages because of positive retest effects or other factors, at least in some respects longitudinal changes may have nearly the same meaning across all of adulthood.

Keywords

aging; cognitive change; longitudinal; reliability

A substantial discrepancy exists between cross-sectional and longitudinal age trends in neuropsychological and cognitive functioning because significant cross-sectional age differences are apparent beginning when adults are in their 20s and 30s, but there is seldom significant longitudinal decline until adults are 60 years of age or older (e.g., Ronnlund, Nyberg, Backman, & Nilsson, 2005; Salthouse, 2009; Schaie, 2005). One possible implication of these results is that any changes that occur in early adulthood reflect random variation, and do not represent the same developmental phenomenon, or have the same meaning, as changes occurring later in adulthood. The primary goal of the current report was to investigate this implication by examining the nature of neuropsychological and cognitive change at different periods in adulthood. The rationale was that if true maturational change occurs only in late adulthood, then the measures of change in young adults should have weak reliability and correlations among the change measures should be close to zero. That is, if measures of change among young adults primarily reflect random variation, they would not be expected to be very consistent or reliable.

However, an alternative interpretation is that longitudinal change represents a mixture of influences, including not only maturation but also practice effects associated with prior testing experience. According to this perspective, systematic neurocognitive change could occur continuously throughout adulthood, but mean declines may not be evident at younger ages because of the presence of large practice effects (e.g., Salthouse, 2009; Salthouse &

Correspondence concerning this article should be addressed to Timothy A. Salthouse, 102 Gilmer Hall, Department of Psychology, University of Virginia, Charlottesville, VA 22904-4400. salthouse@virginia.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Tucker-Drob, 2008). These two interpretations therefore differ in their predictions about the relationship between age and the reliability of measures of longitudinal change. That is, if longitudinal change primarily reflects random variation at young ages but more systematic effects of maturational decline at older ages, then one would expect estimates of the reliability of the longitudinal changes to be larger at older ages. In contrast, if the changes are just as systematic at young ages as at old ages, but mainly differ in the mean value because of differential retest influences or other factors, then estimates of the reliability of change in young adulthood should be similar in magnitude to estimates of the reliability of change in later adulthood. Unfortunately, few studies have reported information about the reliability of longitudinal change, and apparently none have examined reliability of change in neuropsychological and cognitive variables as a function of age.

Simple change or difference scores for a single variable at two occasions often have low reliability, largely because the correlation between the scores at the two occasions (i.e., the stability coefficient) is frequently high relative to the reliability of scores at each occasion. That is, the expected reliability of a difference between scores at Time 1 and Time 2 can be predicted from the equation:

$$\text{Reliability of diff}_{12}=[(\text{rel}_1+\text{rel}_2)/2-\text{corr}_{12}]/1-\text{corr}_{12}, \quad (1)$$

and therefore "as the correlation between the two variables approaches their average reliability, the reliability of the difference score approaches zero (Cohen & Cohen, 1983, p. 69)."

Inspection of equation (1) suggests that there are at least two ways of increasing the reliability of a change score; reducing the magnitude of the correlation across occasions (i.e., corr_{12}), or increasing the reliability of the measurements at each occasion (i.e., rel_1 and rel_2). Although the across-occasion correlation might be expected to decrease with increases in the length of the interval between occasions, it cannot be manipulated directly. In order to optimize reliability of changes across relatively short intervals, therefore, the most practical approach is to try to increase the reliability of the measurements at each occasion. Reliability can generally be increased by expanding the number of relevant observations, either by aggregation across multiple assessments of the same variable with different versions of the test, or by aggregation across different variables representing the same ability.

The current project relies on both of these methods for maximizing reliability at each occasion in an attempt to obtain the most reliable assessments of short-term longitudinal change. The data were derived from a longitudinal study in which three versions of each test were performed on separate sessions at each occasion, and because the tests were selected to represent distinct ability factors (e.g., reasoning, spatial visualization, memory, and speed), additional analyses were conducted on composite scores created by averaging scores from the three variables representing each ability.

Although equation (1) can be used to predict the expected reliability of a longitudinal change score, expected reliability does not necessarily correspond to actual observed reliability. Because all of the data in the current project were recorded at the level of individual items, it was possible to compute direct estimates of change reliability. That is, the availability of item-level data allowed separate scores to be computed for odd-numbered and even-numbered items at each occasion, as well as separate change scores for the two types of items. The two change scores were then used as "items" in the computation of coefficient alpha estimates of the reliability of the longitudinal changes. For example, in a test with four items, separate scores at each occasion would be based on the sum of items 1 and 3, and on the sum of items 2 and 4, and then two across-occasion changes would be derived (i.e., Δ_{1+3} and Δ_{2+4}), which could serve as the input values in the coefficient alpha computation. Because three different versions of each test were performed at each occasion, the scores for odd-numbered items were averaged across the three versions as were the scores for even-numbered items, and then coefficient alphas was determined from the average changes in the averaged odd-item scores and the average changes in the even-item scores.

A decision had to be reached about the best method of examining the relation of age to the reliability of change. Moderately large samples of participants are needed to obtain precise estimates of reliability, which is optimized by dividing the sample into as few groups as possible. However, a large number of different age groups is desirable to evaluate age trends. A compromise between these two conflicting goals was adopted in the current study by conducting analyses on overlapping age groups. That is, 12 overlapping age groups, each spanning a 20-year age range, were created with successive groups differing by 5 years in initial age. This smoothing provides a clearer representation of the age trends, while maintaining moderately large sample sizes for the computation of reliability.

Method

Participants

Characteristics of the 420 participants, arbitrarily divided into three age groups to facilitate description of the sample, are summarized in Table 1. The results reported in the figures are based on mixtures of these groups, created by shifting a 20-year-age-range by 5 years for each successive group.

It can be seen that increased age was associated with somewhat lower levels of self-rated health, but with higher levels of education and age-adjusted Digit Symbol and Logical Memory scaled scores. These latter results suggest that older adults in the sample were more select relative to their age peers in the nationally representative sample used to establish norms for the tests than were the younger adults. The retest interval between occasions was deliberately varied across participants, and ranged from 1 to 4 years with an average of about 2.2 years. However, it is important to note that there was no relation between the age of the participant and the length of the retest interval.

Variables

At each occasion the participants performed 12 neuropsychological or cognitive tests selected to reflect four different abilities. Reasoning ability was represented by Matrix Reasoning, Shipley Abstraction, and Letter Sets; Spatial Visualization ability was represented by Spatial Relations, Paper Folding, and Form Boards; Episodic Memory ability was represented by Word Recall, Paired Associates and Logical Memory; and Perceptual Speed ability by Digit Symbol, Pattern Comparison, and Letter Comparison. The tasks, which are briefly described in the appendix, have been described in more detail in other articles (e.g., Salthouse & Ferrer-Caja, 2003; Salthouse, Pink, & Tucker-Drob, 2008; Salthouse, Siedlecki, & Krueger, 2006). These other articles also report coefficient alpha estimates of the reliabilities, which ranged from .71 to .91, and results of confirmatory factor analyses establishing the construct validity of the variables. Four vocabulary tests were also performed at each occasion, but those data are not reported here because vocabulary scores usually exhibit little or no change across most of adulthood.

There were three different versions of each test, which were performed on separate sessions within a period of approximately two weeks on both occasions. The coefficient alpha reliabilities were generally similar with each test version (cf. Salthouse & Nesselroade, under review).

Results

The first step in the analyses consisted of computing the mean and standard deviation of the scores on each version of each test at the initial occasion. The scores at both occasions were then converted into z-score units based on the distribution at the first occasion. Expressing all scores in the same units facilitated comparisons of the same variable across occasions, and comparisons across different variables.

Changes were initially examined at the level of test scores aggregated across the three assessments (i.e., test versions) at each occasion. The mean longitudinal changes (i.e., score at the second occasion minus score at the first occasion) for the individual variables are portrayed in Figure 1. Notice that for most variables the mean longitudinal changes shift from positive to negative with increasing age. The reasoning variables are an exception to this pattern as these changes were relatively small, and slightly positive at all ages.

Reliabilities of the longitudinal changes were computed by determining the change for the mean of the odd-numbered items across the three test versions and the change for the mean of the even-numbered items across the three test versions, and then treating these two change scores as “items” in coefficient alpha. The reliabilities computed in this manner are portrayed in Figure 2, where it can be seen that most were in the .4 to .6 range, although the Word Recall, Letter Comparison, and Pattern Comparison variables had higher values. Reliability of the changes in the Paper Folding variable were somewhat lower with increased age, whereas the reliability of the changes in the Logical Memory variable were somewhat higher with increased age, but with these exceptions, the magnitudes of most of the reliabilities were nearly constant at all ages.

Similar analyses were also computed with composite scores created from the scores for the three tests representing each cognitive ability. The procedure was identical to that described above except that the scores for the three tests representing each ability were averaged to create composite ability scores. Mean longitudinal changes for the composite ability scores are portrayed in the top left panel of Figure 3. Notice that, similar to the pattern with individual tests in Figure 1, there was a systematic shift in the mean value of longitudinal change from positive to negative with increased age.

Reliabilities of the longitudinal changes in the composite variables are displayed in the top right panel of Figure 3. Two points should be noted about these data. First, the reliabilities were generally higher than those for individual variables, as many were in the .5 to .7 range compared to the .4 to .6 range for individual variables. And second, although there was a slight increase in reliability of the memory composite with increased age, and a decrease in reliability of the spatial visualization composite at the oldest ages, the dominant pattern was one of constancy across adulthood.

Finally, correlations were computed between the changes in the ability composites, which were then adjusted for unreliability by dividing the observed correlation by the square root of the product of the two reliabilities (cf. Cohen & Cohen, 1983, p. 69). These disattenuated correlations are displayed in the bottom panel of Figure 3. It can be seen that although there was a great deal of variability in the correlations, most of them were between .2 and .4. It is important, however, that there is little systematic relation of the disattenuated correlations with age. The correlations between memory and speed changes, and between reasoning and speed changes, were very low at the youngest ages, but the correlations between spatial visualization and memory, between spatial visualization and reasoning, and between reasoning and memory were highest in the youngest ages.

Discussion

There are three major results of this study, which are evident at both the level of individual tests and the level of ability composites. First, for many variables the longitudinal change shifts from positive (reflecting better performance on the second occasion) to negative (reflecting poorer performance on the second occasion) with increased age. The systematic relations between age and the direction and magnitude of longitudinal change may have been neglected in much of the prior longitudinal research because the focus has only been on the age at which the change is significantly negative, rather than on the complete relation between age and change. A key question, which cannot be answered with the current data, is whether the relative contribution of different determinants of cognitive change varies across adulthood. For example, the decreasing age-change relations could reflect diminishing benefits of prior test experience with nearly constant maturational declines, similar retest effects with progressively larger maturational declines, or various combinations of these or other influences. Distinguishing among these possibilities should be a high priority for future research.

The second major result of the current project is that reliabilities of the longitudinal changes were all relatively low, with most of them below about .6. The low values are not particularly surprising because the across-occasion correlations were high. For example, the stability coefficients for scores from single tests ranged from .67 to .91, with a median of .83, and those for the ability composites ranged from .79 to .90, with a median of .87. Nevertheless, it is important to recognize that the reliabilities of changes are typically rather low, because the correlation of a change score with itself, which is one way of thinking of reliability, sets an upper limit on the magnitude of correlations that are possible between the change scores and other variables. Reliability therefore needs to be considered, and possibly adjusted for, when interpreting correlations of change scores.

One method of adjusting correlations for unreliability consists of expressing the correlation as a proportion of the reliable variance, instead of the total variance, by dividing the observed correlation by the average reliability of the scores being correlated. As seen in the bottom of Figure 3, these disattenuated correlations were generally between .2 and .4, suggesting modest correlations among the short-term longitudinal changes in different neurocognitive variables. Other methods could also be used to adjust for unreliability or minimize measurement error, such as latent difference scores or latent growth curve models (e.g., McArdle, 2009), although these more complex methods tend to involve additional assumptions that may not be easily testable.

The third major result of the project was that there was little relation of age to the reliability of the longitudinal changes. Although the longitudinal change shifted in direction from positive to negative with increased age, the reliabilities of the changes were nearly constant across the adult years. It therefore does not appear to be the case that longitudinal change is only reliable when it is negative, at older ages. The correlations of the changes with each other were also generally similar at all ages. Because one way of inferring the meaning of a variable consists of examining its relations to other variables, these results suggest that, at least in some respects, the measures of neurocognitive change appear to have similar meaning at different periods in adulthood.

The current results indicate that longitudinal change is not very reliable across intervals of 1 to 4 years. It is possible that the absolute level of reliability might be higher with longer intervals between test occasions, and perhaps even differentially greater at older ages, such that age differences in reliability of change emerge. However, there is no evidence that change is less reliable at younger ages than at older ages across the short intervals in the current project. Despite different magnitudes and directions of longitudinal change at different ages, the changes appear to be just as systematic in early and middle adulthood as in late adulthood.

In conclusion, the results of this project suggest that longitudinal change in neuropsychological and cognitive variables occurs continuously from early adulthood, and that even though the direction of change shifts with increasing age, some of the properties of change are very similar across all of the adult years. The mean value of longitudinal change is often positive among young adults, possibly because of large benefits of retest effects, but there was little relation of age to the reliabilities of the changes, or to the correlations of the changes with each other, which suggests that in certain respects the measures of cognitive change may have the same meaning at different ages.

Acknowledgments

This research was supported by National Institute on Aging Grant NIA R37AG024270.

References

Bennett, GK.; Seashore, HG.; Wesman, AG. Differential Aptitude Test. San Antonio, TX: The Psychological Corporation; 1997.

Cohen, J.; Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. 2. Hillsdale, NJ: Erlbaum; 1983.

Ekstrom, RB.; French, JW.; Harman, HH.; Dermen, D. Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service; 1976.

McArdle JJ. Latent variable modeling of differences and changes with longitudinal data. Annual Review of Psychology. 2009; 60:577–605.

Raven, J. Advanced progressive matrices, set II. London: H. K. Lewis; 1962.

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

Salthouse TA, Babcock RL. Decomposing adult age differences in working memory. Developmental Psychology. 1991; 27:763–776.

Salthouse TA, Ferrer-Caja E. What needs to be explained to account for age-related effects on multiple cognitive variables? Psychology and Aging. 2003; 18:91–110. [PubMed: 12641315]

Salthouse TA, Fristoe N, Rhee SH. How localized are age-related effects on neuropsychological measures? Neuropsychology. 1996; 10:272–285.

Salthouse TA, Pink JE, Tucker-Drob EM. Contextual analysis of fluid intelligence. Intelligence. 2008; 36:464–486. [PubMed: 19137074]

Salthouse TA, Siedlecki KL, Krueger LE. An individual differences analysis of memory control. Journal of Memory and Language. 2006; 55:102–125.

Salthouse TA, Tucker-Drob EM. Implications of short-term retest effects for the interpretation of longitudinal change. Neuropsychology. 2008; 22:800–811. [PubMed: 18999354]

Salthouse, TA.; Nesselroade, JR. Dealing with short-term fluctuation in longitudinal research. n.d. Manuscript under review

Schaie, KW. Developmental influences on adult intelligence: The Seattle Longitudinal Study. New York: Oxford University Press; 2005.

Wechsler, D. Wechsler Adult Intelligence Scale. 3. San Antonio, TX: The Psychological Corporation; 1997a.

Wechsler, D. Wechsler Memory Scale. 3. San Antonio, TX: Psychological Corporation; 1997b.

Zachary, RA. Shipley Institute of Living Scale–Revised. Los Angeles: Western Psychological Services; 1986.

Appendix. Description of Reference Variables and Sources of Tasks

Variable	Description	Source
Matrix Reasoning	Determine which pattern best completes the missing cell in a matrix	Raven (1962)
Shipley Abstraction	Determine the words or numbers that are the best continuation of a sequence	Zachary (1986)
Letter Sets	Identify which of five groups of letters is different from the others	Ekstrom, et al. (1976)
Spatial Relations	Determine the correspondence between a 3-D figure and alternative 2-D figures	Bennett, et al. (1997)
Paper Folding	Determine the pattern of holes that would result from a sequence of folds and a and a punch through folded paper	Ekstrom, et al. (1976)
Form Boards	Determine which combinations of shapes are needed to fill a larger shape	Ekstrom, et al. (1976)
Logical Memory	Number of idea units recalled across three stories	Wechsler (1997b)
Free Recall	Number of words recalled across trials 1 to 4 of a word list	Wechsler (1997b)
Paired Associates	Number of response terms recalled when presented with a stimulus term	Salthouse, et al. (1996)
Digit Symbol	Use a code table to write the correct symbol below each digit	Wechsler (1997a)
Letter Comparison	Same/different comparison of pairs of letter strings	Salthouse & Babcock (1991)
Pattern Comparison	Same/different comparison of pairs of line patterns	Salthouse & Babcock (1991)

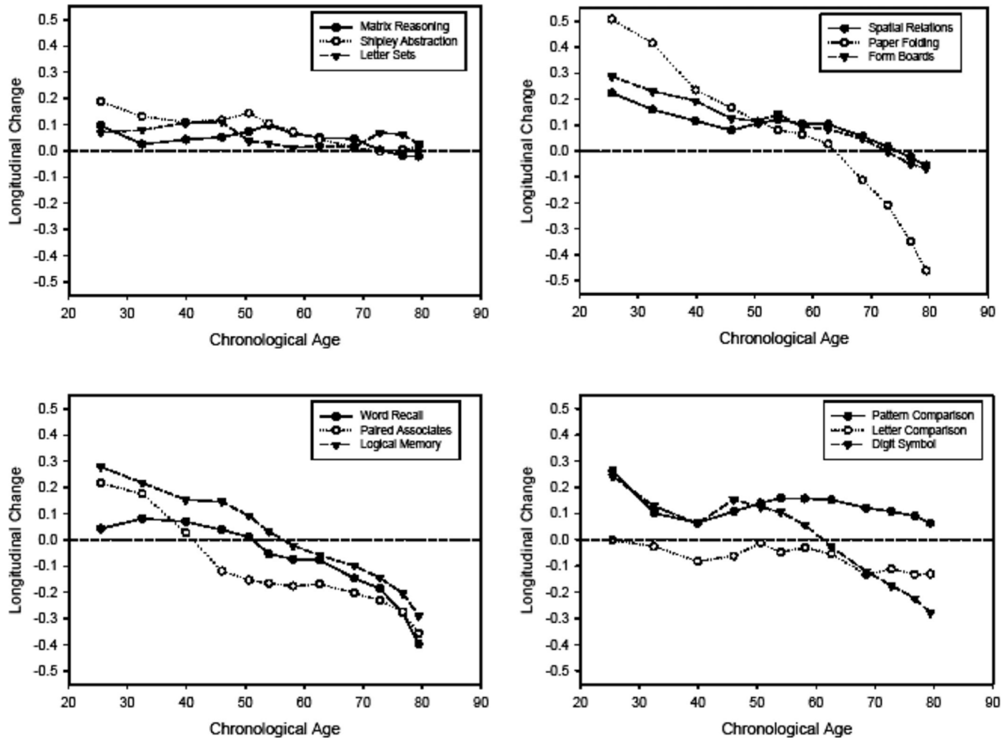

Figure 1.

Mean longitudinal change (second score minus first score) in individual variables in first-occasion z-score units for 12 overlapping age groups.

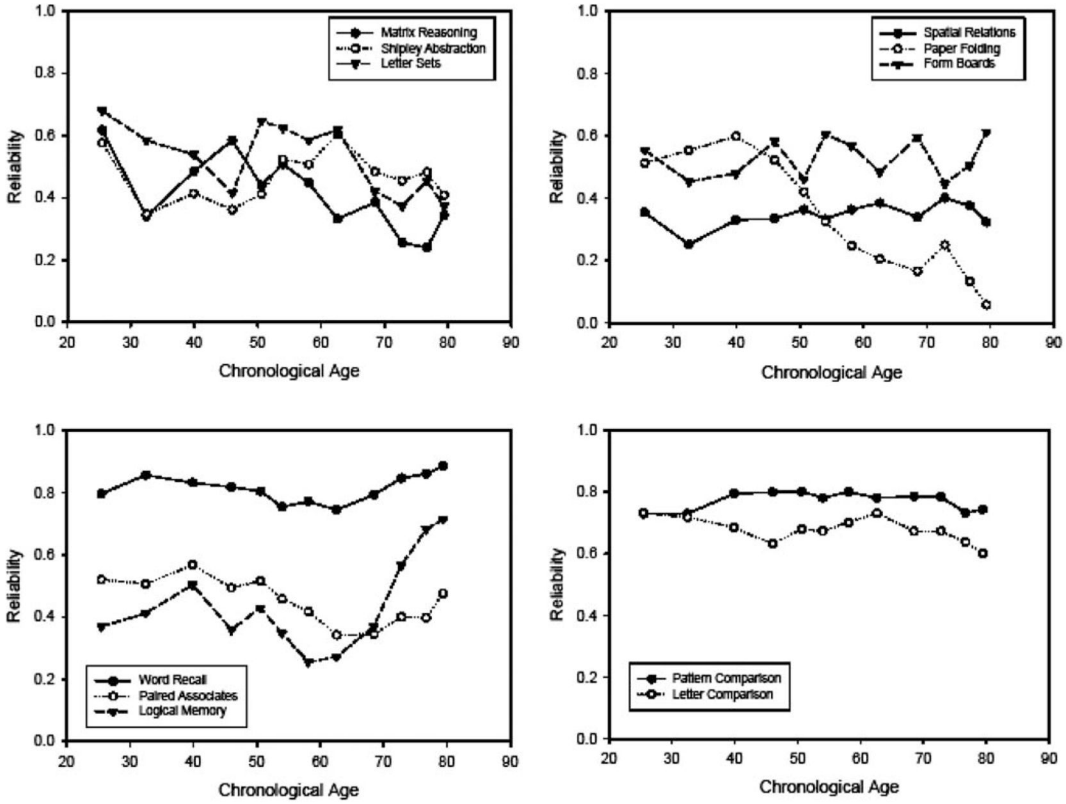

Figure 2.

Coefficient alpha estimates of the reliability of longitudinal changes in individual variables for 12 overlapping age groups.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

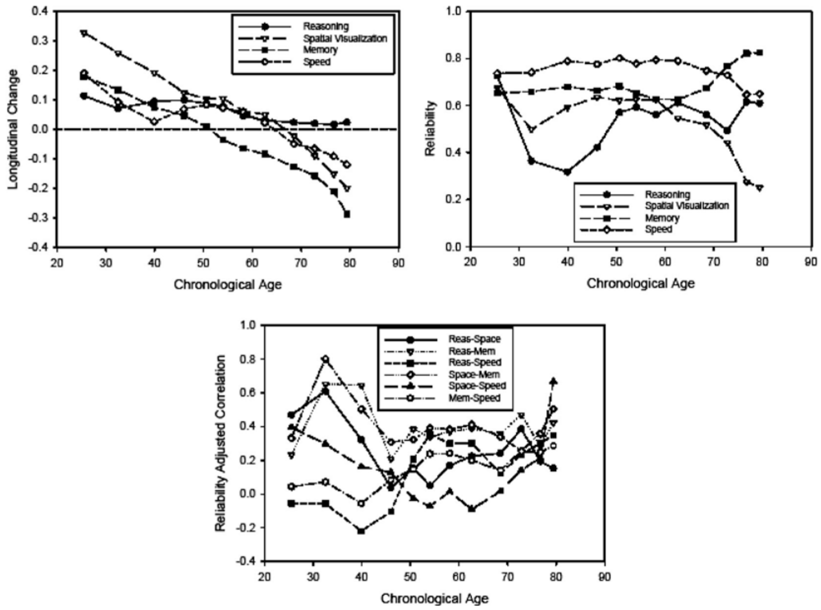

Figure 3.

(A) Mean longitudinal change in first-occasion z-score units for composite ability scores for 12 overlapping age groups. (B) Coefficient alpha estimates of the reliability of longitudinal changes in composite ability scores for 12 overlapping age groups. (C) Reliability-adjusted correlations between changes in composite scores for 12 overlapping age groups.

Table 1

	Age group			
	18–39	**40–59**	**60–95**	**Age r**
N	89	152	179	
Age at Time 1	26.0 (6.3)	51.3 (5.3)	72.0 (7.6)	
Prop. Females	.53	.74	.59	−.01
Health	2.3 (0.9)	2.2 (0.8)	2.4 (0.9)	.14*
Education	14.8 (2.2)	15.8 (2.4)	16.1 (2.9)	.21*
Scaled Scores				
Vocabulary	12.8 (3.1)	12.6 (2.9)	13.5 (2.7)	.11
Digit symbol	11.1 (2.7)	11.6 (3.0)	11.9 (2.6)	.18*
Logical memory	11.3 (2.7)	11.7 (2.8)	12.6 (2.6)	.21*
Word recall	12.1 (3.6)	12.5 (3.6)	12.8 (3.0)	.07
Retest interval (years)	2.2 (0.6)	2.1 (0.6)	2.2 (0.5)	.00

Note. Health is a rating on a 5-point scale ranging from 1 for excellent to 5 for poor. Scaled scores are age-adjusted scores which have means of 10 and standard deviations of 3 in the nationally representative normative samples (i.e., Wechsler, 1997a, 1997b).

* $p < .01$.