



Published in final edited form as:

Psychol Assess. 2014 December ; 26(4): 1063–1069. doi:10.1037/pas0000007.

Frequent assessments may obscure cognitive decline

Timothy A. Salthouse

Department of Psychology, University of Virginia.

Abstract

Effects of an initial testing experience on the level of cognitive performance at a second occasion are well documented. However, less is known about the effects of additional testing experiences beyond the first on the amount of cognitive change over a specified interval. This issue was investigated in a moderately large sample of adults between 18 and 95 years of age who performed a battery of cognitive tests either two or three times at variable intervals between each assessment. Multiple regression analyses were used to examine effects of the number of assessments on change while controlling the length of the interval between the first and last assessments. Change in each of five cognitive domains was less negative when there was an intervening assessment. To illustrate, for adults between 65 and 95 years of age, the estimated change from a first to a second assessment across an average interval of 3.9 years was $-.25$ standard deviation units ($p < .01$), but it was only $-.06$ standard deviation units, and not significantly different from zero, when an intervening assessment occurred during the interval. These results indicate that cognitive change may not be detected when individuals are assessed frequently with relatively short intervals between the assessments.

Keywords

longitudinal; neuropsychological change; serial assessments; retest effects; age comparisons; early detection of change

Direct measurement of change requires a minimum of two assessments, but precision in characterizing the trajectory of change is enhanced with additional assessments. Although more information is generally desirable, the primary question in the current study was whether the estimates of change over a given interval differ when an additional assessment occurs between the measurements of primary interest. That is, do estimates of the magnitude of cognitive change vary according to the number of assessments occurring during the interval? Secondary questions are if an intervening assessment affects the magnitude of change, does the timing of the additional assessment matter, and does the effect of an additional assessment, and the timing of that assessment, vary according to the individual's age? Greater benefit of an intervening assessment might be expected if it occurs later in the interval after a greater amount of change has occurred, and older individuals might benefit more from an additional assessment than younger individuals if they have experienced more negative change by the time of the intervening assessment.

Successive assessments of cognitive functioning are often obtained after relatively short intervals to maximize sensitivity to detect change, but it is possible that frequent assessments have the opposite effect of obscuring change if the process of measurement alters the magnitude of change. That is, scores on a second assessment are often higher than those on the initial assessment because of practice effects (Calamia, Markon & Tranel, 2012; Hausknecht, Halport, Di Paolo & Moriarty Gerrard, 2007), and if the practice effects have not dissipated by the time of the next assessment, they could obscure any decline that may be occurring.

The issue of the number of assessments on change also has clinical implications because the magnitude of change over a given interval may have different meaning according to the number of assessments between the first and last measurement. For example, an individual may be considered to have remained stable over a 3-year interval if tests are administered every year, whereas appreciable decline might be detected if assessments were only obtained at the beginning and end of the interval.

Because the intervals between successive assessments are typically constant in most longitudinal studies, it has been difficult to investigate effects of the number of intervening assessments on cognitive change while controlling the length of the interval between relevant assessments. Participants in longitudinal studies are sometimes reported to have missed scheduled assessments (e.g., Colsher & Wallace, 1991; Fairias, Cahn-Weiner, Harvey, Reed, Mungas, et al., 2009; Gow, Corley, Starr & Deary, 2012; Granholm, Link, Fish, Kraemer & Jeste, 2010; Hayden, Reed, Manly, Tommet, Pietrzak, et al., 2011), but there have apparently not been any analyses of the effects of different numbers of assessments on measures of change across the same interval.

However, the effect of an additional assessment on cognitive change can be investigated with data from the Virginia Cognitive Aging Project (VCAP; Salthouse, 2009; Salthouse, in press-c; Salthouse, Pink & Tucker-Drob, 2008) because the intervals between successive assessments in that project were deliberately varied across participants, and a moderately large number of participants have completed either two or three longitudinal assessments with tests evaluating several major cognitive domains. These characteristics make it possible to compare the magnitude of change in different domains over the same interval in individuals who have completed either two, or three, assessments.

Some information relevant to the current questions is available in prior reports of the VCAP study. For example, Salthouse (2011) examined the relation of change to the interval between the first and second assessments in a subset of participants from the current sample (i.e., 1,576 of the 2,082). The major finding in that report was that change in several abilities was more negative with longer intervals between assessments. However, only data across two assessments were considered in that report, and thus no information was available about the effect of number of assessments on change.

Another report (Salthouse, 2013b) capitalized on the fact that some participants in VCAP performed different tests on the second and third sessions of the first occasion instead of parallel versions of the same tests. That is, the VCAP study involves a measurement burst

design in which individuals participate in three sessions within a period of approximately two weeks at each occasion. Analyses in that study revealed that additional experience with parallel versions of the same tests was associated with more positive longitudinal change over an interval of about 2.5 years than comparable amounts of experience with different types of tests. Although change was more positive among participants with more relevant experience, all of the experience occurred at the first occasion, and was not distributed across different occasions as was the case in the present study.

To summarize, the primary goal of the present study was to determine the effect of an intervening assessment on the longitudinal change in different cognitive domains. The interval between successive assessments varied across participants, and thus it was possible to compare change over the same average interval for participants with either two or three assessments. Because the participants varied from 18 to 95 years of age, separate analyses were carried out in each of three age groups in addition to analyses on the complete sample.

Methods

Participants

The data were based on VCAP participants who had completed either two or three longitudinal assessments. The sample was divided into three age groups with participants between 18 and 39 years of age in one group, those between 40 and 64 years of age in a second group, and those between 65 and 95 years of age in a third group. The research was approved by the local Institutional Review Board.

Characteristics of the participants according to age group and number of assessments are reported in Table 1. It can be seen that the proportion of females was highest in the younger group, but that participants in the older groups had a greater number of years of education, higher estimated IQs (see below), and poorer self-rated health. The interval between the first and the last (either the second or the third) assessment was greater for participants with three assessments compared to those with two, but there was moderate variability in the intervals among participants with both two and three assessments.

There were a variety of reasons for differences in the number of assessments, including greater opportunity for more assessments when the initial assessment was early in the history of the project. Analyses of variance on the composite scores (see below) were conducted to investigate possible differences at the first measurement occasion between participants with two or three assessments. The results revealed a significant effect of age group in every composite score, significant effects of the number of assessments with the memory and speed composite scores, and a significant interaction of age and number of assessments in every cognitive domain. The age effects indicate that performance was higher at younger ages in all composite scores except vocabulary, and the effects of number of assessments indicate that participants with three assessments generally had higher levels of performance at the initial occasion than participants with two assessments. The interactions reflect larger differences between individuals with two versus three assessments in the older groups compared to the younger group.

Assessment of sample representativeness

Because the participants in VCAP reflect a convenience sample, it is important to characterize the sample relative to a broader population. In a recent study (Salthouse, in press-a) both the VCAP test battery and the Wechsler Adult Intelligence Scale IV (Wechsler, 2008) test battery were administered to 90 adults between 20 and 80 years of age, which allowed estimates of full scale IQ scores to be derived in VCAP participants. Because IQ scores are age-adjusted, the estimation procedure consisted of partialling age from the raw scores to create residual scores, determining the best prediction of IQ from the residual scores, and then using the resulting regression equation to estimate IQ. The most parsimonious regression equation with good prediction of IQ (i.e., $R^2 = .86$) was: $109.32 + 2.47$ (series completion residual) $+ 1.54$ (antonym vocabulary residual) $+ 1.78$ (paper folding residual). This equation was applied to all of the VCAP participants with relevant data to generate estimated IQ values. IQs in the nationally representative normative sample have a mean of 100 and a standard deviation of 15 (Wechsler, 2008). Because the mean IQs in Table 1 range from about 105 to 113, with standard deviations between 12 and 15, the participants in the present sample can be inferred to have a higher average level of functioning than the normative sample, but approximately the same degree of variability.

Cognitive functioning

The tests in VCAP were selected to represent broad dimensions of cognitive functioning, including ability domains that exhibit early age-related declines, such as speed and memory, and domains such as word knowledge that tend to be maintained until late life. The 16 cognitive tests, and their reliabilities and validities, have been described in other publications (Salthouse, 2009; Salthouse, in press-c; Salthouse et al., 2008), and thus they are only briefly described here. Episodic memory was assessed with the Logical Memory test from the Wechsler Memory Scale III (Wechsler, 1997b), the Word List Test from the Wechsler Memory Scale III (Wechsler, 1997b), and a locally developed Paired Associates test (Salthouse, Fristoe & Rhee, 1996). Speed was measured with Digit Symbol (Wechsler, 1997a), Letter Comparison (Salthouse & Babcock, 1991), and Pattern Comparison (Salthouse & Babcock, 1991) tests. Vocabulary was measured with WAIS III Vocabulary (Wechsler, 1997a), Picture Vocabulary from the Woodcock-Johnson Cognitive Ability test (Woodcock & Johnson, 1989), Antonym Vocabulary (Salthouse, 1993), and Synonym Vocabulary tests (Salthouse, 1993). Reasoning was assessed with the Raven's Advanced Progressive Matrices (Raven, 1962), Shipley Abstraction (Zachary, 1986), and Letter Sets (Ekstrom, French, Harman, & Dermen, 1976) tests. Spatial visualization was assessed with the Spatial Relations test from the Differential Aptitude Test Battery (Bennett, Seashore & Wesman, 1997), the Paper Folding test from the Educational Testing Service Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976), and the Form Boards test (Ekstrom et al., 1976).

Scores in each test were converted to *z*-scores based on the mean and standard deviations of the complete sample at the first assessment, and composite scores formed for each ability domain by averaging *z*-scores for the relevant tests. Composite scores were selected because they are more reliable than scores of individual tests, and may better represent the relevant ability because test-specific influences are averaged out when forming the composites.

Coefficient alphas for the composite scores based on the intercorrelations of the tests representing each ability domain were .78 for episodic memory, .83 for perceptual speed, .91 for vocabulary, .84 for reasoning, and .83 for spatial visualization.

A measure of general cognitive ability was obtained from the first principal component (PC1) in a principal components analysis of the 16 tests at the first occasion in the entire sample. The PC1 was associated with 42.7% of the variance in the test scores, and it had correlations with the composite scores of .71 for memory, .65 for speed, .66 for vocabulary, .91 for reasoning, and .82 for spatial visualization, and a correlation of .84 with estimated IQ.

Measurement burst design

Robustness of the effects was examined across scores on different versions of the tests that were administered in separate sessions in a measurement burst design. The measurement burst design implemented in VCAP involved participants performing different versions of the 16 tests on each of three sessions completed within a period of about two weeks at each occasion. Possible differences in mean performance across versions were adjusted with regression equations derived from data of a sample of participants who performed the three versions in counterbalanced order (Salthouse, 2007). Some of the participants performed different types of tests on the second and third sessions (Salthouse, 2013b), and therefore the total sample sizes were 2,263 for session 1, but only 1,060 for sessions 2 and 3.

Results

The initial analyses were analyses of variance on the scores at the first and final (i.e., either 2nd or 3rd) occasion, with age group (18-39, 40-64, or 65-95) and number of assessments (2 or 3) as between-subjects factors and time (first or final assessment) as a within-subjects factor. Because participants with two or three assessments differed in their level of performance at the first occasion and in the length of the interval between the first and final assessment (cf. Table 1), two covariates were used to control these differences when examining change. One covariate was the PC1 as an estimate of general cognitive ability, and the other was the interval between the first and final assessment. The effect of these covariates was to conduct the analyses at the average PC1 value and the average interval between the first and final assessments.

Because there were many more participants with data on session 1 than on sessions 2 and 3, one set of analyses examined only data from session 1, and a second set examined data from all three sessions. Results from both sets of analyses are reported in Table 2, with the top panel containing the results with only session 1 data, and the bottom panel containing the results with data from all three sessions.

The main effects of age in each analysis reflect the higher performance at younger ages, except for vocabulary where performance was higher at older ages. The main effects of time indicate that the level of performance differed between the first and the final assessment, and the interactions of age and time indicate that the time-related differences were more negative at older ages. The main effects of number of assessments for memory and speed abilities

indicate that, when collapsed across time and age, performance was higher with three assessments than with two assessments. Of greatest interest for the present purpose are the interactions of number of assessments and time which indicate that for all abilities except reasoning, change was less negative when there were three assessments compared to when there were only two assessments. However, none of the interactions of age with number of assessments and with time were significant, and therefore there was no evidence that the benefits of an additional assessment varied across age groups.

The analyses based on data from participants with all three sessions included session as an additional factor in the analyses. The results were similar to the analyses with only session 1 data in terms of the significant age and time main effects and age-by-time interactions, and the significant number of assessments-by-time interaction for memory, vocabulary and reasoning, indicating less negative change with three compared to two assessments. Importantly, none of the interactions of session were significant, and therefore there was no evidence that the pattern of less negative change with an intervening assessment varied across sessions.

The session effects are similar to those reported with analyses of subsets of these data in Salthouse (2012; 2013a). That is, the means were higher on later sessions, with greater across-session increases in younger adults for vocabulary, but greater increases in older adults for spatial visualization.

Figure 1 portrays the estimated memory composite scores (means and standard errors) for the first session at each measurement occasion for participants in three age groups after statistical control of the length of the T1 to T_n interval and the PC1 measure of general cognitive ability. Notice that in each group the change from the first to the last occasion was more positive with three assessments than with two assessments. The results in the oldest group are particularly noteworthy because significant decline was only evident in participants without an additional assessment during the longitudinal interval.

Estimates of the T_n – T1 composite score differences on session 1 for participants with two or three assessments were computed in each cognitive domain after statistical control of the length of the total interval and the PC1 measure of general cognitive ability. These values are portrayed in Figure 2, where it can be seen that although there was variability in the absolute values of change across domains, in each case the changes were more positive with three assessments (open symbols) than with two assessments (filled symbols). Many of the positive changes in Figure 2 are likely attributable to practice effects associated with prior experience with the tests (Salthouse, 2010).

In order to investigate whether the effects of an additional assessment were specific to high-functioning adults, the sample of older adults was divided into two groups based on the MMSE (Folstein, Folstein & McHugh, 1975) score at the second occasion. The high group (N = 328) had MMSE scores between 28 and 30 (mean = 29.1) and the low group (N = 122) had MMSE scores between 23 and 27 (mean = 25.9). The covariate-adjusted T_n-T1 differences in memory were -.22 and -.04 for the participants in the high group with two and three assessments, respectively, and -.32 and -.15 for participants in the low group with

two and three assessments, respectively. These results therefore suggest that even individuals who might be considered at risk for dementia because their MMSE scores were less than 28 exhibited only half as much decline over an interval of about 3.9 years if they were assessed three times instead of only twice.

Finally, effects of the timing of the intervening assessment was investigated in participants with three assessments by examining relations with a measure of the proportion of the total T1-T3 interval occupied by the interval from the 1st (T1) to the 2nd (T2) assessment. As an example of the computation, if the T2 occasion occurred 3 years after T1, and the T3 occasion occurred 2 years after T2, the proportion would be 3/5 or .6. The proportions ranged from .05 to .96, with a mean of .46 and a standard deviation of .15. The correlation of the proportion with age was only .02, and correlations with the T1-T3 differences in the abilities ranged between .00 and .09, with only the correlation with memory (i.e., .09) significantly different from 0. These results suggest that, with the exception of a slightly greater benefit with a longer interval from the first assessment for memory, there were minimal effects on cognitive change of when the intervening assessment occurred.

Discussion

Because they are designed to evaluate change within the same individuals, longitudinal studies necessarily involve repeated assessments. Furthermore, multiple assessments beyond the minimum of two are often considered desirable to increase sensitivity in detecting change. However, a possible disadvantage of frequent assessments is that the phenomenon under investigation could be distorted if the additional assessments are reactive. In fact, the results of this study indicate that estimates of longitudinal change in memory and other cognitive domains are affected by an intervening assessment. For example, the results in Figures 1 and 2 indicate that for individuals 65 years and older, the decline in memory over an interval of almost 4 years would not be significant if an additional assessment occurred during the interval, whereas a significant decline of about .25 standard deviation units would have been detected without an intervening assessment. The large effects of an additional assessment in the change in memory are particularly noteworthy because memory is the cognitive domain most sensitive to dementia and other pathologies (Backman et al., 2005).

The phenomenon of selective attrition refers to the finding that people who return for additional occasions frequently have higher scores at the initial occasion than participants who do not return, and it is important to consider whether an analogous phenomenon might be operating in the current study. There are two reasons why this seems unlikely. First, analyses reported in Salthouse (in press-b) revealed that the observed changes for returning VCAP participants were similar to the imputed changes of participants with only one occasion. These results suggest that although the people who do not return for subsequent occasions may have somewhat lower levels of functioning at the initial occasion than people who do return, the change that they would have exhibited had they returned appears comparable to that of returning participants. In other words, selective attrition is primarily associated with level of functioning and not change in functioning, which is the primary outcome of interest here. A second reason why ability differences seem unlikely to be contributing to the effects of two versus three assessments is that the analyses controlled a

measure of general cognitive ability (i.e., the first principal component), which served to adjust for initial differences between individuals with two and three assessments.

The measurement burst design allowed the robustness of the effects to be examined across multiple sessions. The major finding was that the overall pattern of results was similar in the analyses of all three sessions and in the analysis restricted to session 1. Importantly, the absence of interactions of session with number of assessments and time provides no evidence that the more positive change with an additional assessment varied as a function of session.

An important implication of these results is that frequent assessments could be obscuring cognitive change that might have been detected had there been no intervening assessments. It is therefore possible that the lack of significant decline sometimes reported in studies with annual administrations of identical tests (Johnson et al., 2012; McCleary et al., 1996; Storandt et al., 2002) is at least partially attributable to the positive effects of frequent assessments obscuring decline that would have been detected with fewer assessments. Clinicians interested in optimizing their evaluation of change therefore need to consider whether the greater sensitivity in detecting when change occurs that is achieved by frequent assessments offsets the possibility of obscuring the detection of change because of the reactive effects of each assessment.

The results of this study also have implications for statistical analyses that combine data from participants with different numbers of assessments across the same interval. That is, because the magnitude of change varies according to the number of assessments, estimates of change may be imprecise when there is a non-monotone or intermittent pattern of missing data, and the numbers of assessments are not considered in the analyses.

It is worth considering how future research might be designed to deal with the phenomenon that estimates of change are affected by the number of assessments occurring within a given interval. One option might be to decrease the frequency of assessments for individuals not considered at risk for cognitive decline. Although this may be the simplest solution, it could have the undesirable consequence of reducing sensitivity to detect cognitive change. Another possibility is to administer different tests of the same abilities on successive occasions. For example, story memory tests might be administered on one occasion, and word recall tests on another occasion. The rationale is that the same ability might be evaluated with alternative tests involving different items and requiring somewhat different strategies. Reactive effects might therefore be minimized if those influences are primarily attributable to effects associated with memory for specific test items, or acquisition of test-specific skills and strategies. Future research is needed to determine whether assessments involving different tests that represent the same ability also affect the magnitude of change, but varying the nature of the tests on alternating assessments could be a promising approach to minimize reactive effects.

As with all research, this study has a number of limitations. First, most of the participants were healthy and relatively high functioning, and effects of an additional assessment on change might not be evident to the same extent in clinical groups. Second, the intervening

assessment in this project was identical to the first and last assessments, and very little is known about how similar the intervening experiences must be to have an effect on change. And third, only the effects of a single intervening assessment were examined, and it is not known whether effects of additional assessments on change accumulate such that even moderate decline might not be detected when assessments are repeated at relatively short intervals.

Despite these limitations, the effects of intervening assessments on cognitive change can be considered robust because they were apparent in different ability domains and across a wide range of ages. It is therefore important to recognize that longitudinal research involves tradeoffs because although multiple assessments are clearly desirable to provide the most accurate characterization of the developmental trajectory, each assessment has the potential to distort the phenomenon under investigation. Indeed, the results of this study indicate that frequent assessment with identical tests may obscure decline in cognitive functioning.

Acknowledgments

This research was supported by NIH Grant R37AG024270.

References

- Backman L, Jones S, Berger A-K, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's Disease: A meta-analysis. *Neuropsychology*. 2005; 19:520–531. [PubMed: 16060827]
- Bennett, GK.; Seashore, HG.; Wesman, AG. *Differential Aptitude Test*. Psychological Corporation; San Antonio, TX: 1997.
- Calamia M, Markon K, Tranel D. Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*. 2012; 26:543–570. [PubMed: 22540222]
- Colsher PL, Wallace RB. Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*. 1991; 1:215–230. [PubMed: 1669503]
- Fairias ST, Cahn-Weiner DA, Harvey DJ, Reed BR, Mungas D, Kramer JH, Chui H. Longitudinal changes in memory and executive functioning are associated with longitudinal change in instrumental activities of daily living in older adults. *The Clinical Neuropsychologist*. 2009; 23:446–461. [PubMed: 18821181]
- Ekstrom, RB.; French, JW.; Harman, HH.; Dermen, D. *Manual for Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service; Princeton, NJ: 1976.
- Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198.
- Gow AJ, Corley J, Starr JM, Deary IJ. Reverse causation in activity-cognitive ability associations: The Lothian Birth Cohort, 1936. *Psychology and Aging*. 2012; 27:250–255. [PubMed: 21644808]
- Granhölm E, Link P, Fish S, Kraemer H, Jeste D. Age-related practice effects across longitudinal neuropsychological assessments in older people with schizophrenia. *Neuropsychology*. 2010; 24:616–624. [PubMed: 20804250]
- Hayden KM, Reed BR, Manly JJ, Tommet D, Pietrzak RH, Chelune GJ, Yang FM, Revell AJ, Bennett DA, Jones RN. Cognitive decline in the elderly: An analysis of population heterogeneity. *Age and Ageing*. 2011; 40:685–689.
- Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty Gerrard MO. Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*. 2007; 92:373–385. [PubMed: 17371085]

- Johnson JK, Gross AL, Pa J, McLaren DG, Park LQ, Manly JJ. Longitudinal change in neuropsychological performance using latent growth models: A study of mild cognitive impairment. *Brain Imaging and Behavior*. 2012; 6:540–550. [PubMed: 22562439]
- McCleary R, Dick MB, Buckwalter G, Henderson V, Shankle WR. Full-information models for multiple psychometric tests: Annualized rates of change in normal aging and dementia. *Alzheimer Disease and Associated Disorders*. 1996; 10:216–223. [PubMed: 8939281]
- Raven, J. *Advanced Progressive Matrices, Set II*. H. K. Lewis; London, England: 1962.
- Salthouse TA. Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences*. 1993; 48:P29–P36.
- Salthouse TA. Implications of within-person variability in cognitive and neuropsychological functioning on the interpretation of change. *Neuropsychology*. 2007; 21:401–411. [PubMed: 17605573]
- 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. Effects of age on time-dependent cognitive change. *Psychological Science*. 2011; 22:682–688. [PubMed: 21467547]
- Salthouse TA. Robust cognitive change. *Journal of the International Neuropsychological Society*. 2012; 18:749–756. [PubMed: 22595019]
- Salthouse TA. Effects of age and ability on components of cognitive change. *Intelligence*. 2013a; 41:501–511. [PubMed: 24159248]
- Salthouse TA. Effects of first occasion test experience on longitudinal cognitive change. *Developmental Psychology*. 2013b; 49:2172–2178. [PubMed: 23437801]
- Salthouse TA. Evaluating the correspondence of different cognitive batteries. *Assessment*. (in press-a).
- Salthouse TA. Selectivity of attrition in longitudinal studies of cognitive functioning. *Journal of Gerontology: Psychological Sciences*. (in press-b).
- Salthouse TA. Correlates of cognitive change. *Journal of Experimental Psychology: General*. (in press-c).
- Salthouse TA, Babcock RL. Decomposing adult age differences in working memory. *Developmental Psychology*. 1991; 27:763–776.
- 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]
- Storandt M, Grant EA, Miller JP, Morris JC. Rates of progression in mild cognitive impairment and early Alzheimer's disease. *Neurology*. 2002; 59:1034–1041. [PubMed: 12370458]
- Wechsler, D. *Wechsler Adult Intelligence Scale—Third Edition*. The Psychological Corporation; San Antonio, TX: 1997a.
- Wechsler, D. *Wechsler Memory Scale—Third Edition*. The Psychological Corporation; San Antonio, TX: 1997b.
- Wechsler, D. *WAIS-IV: Administration and scoring manual*. Pearson; San Antonio, TX: 2008.
- Woodcock, RW.; Johnson, MB. *Woodcock–Johnson Psycho-Educational Battery—Revised*. DLM; Allen, TX: 1989.
- Zachary, RA. *Shipley Institute of Living Scale—Revised*. Western Psychological Services; Los Angeles, CA: 1986.

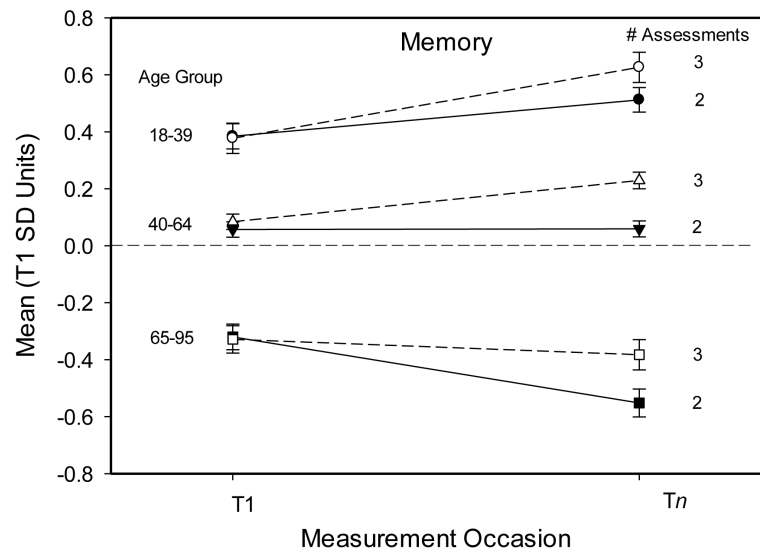


Figure 1. Estimated mean composite memory scores (and standard errors) at the first (T1) and last (Tn) occasion for participants in three age groups with two or three assessments after control of the T1-Tn interval and a measure of general cognitive ability (PC1).

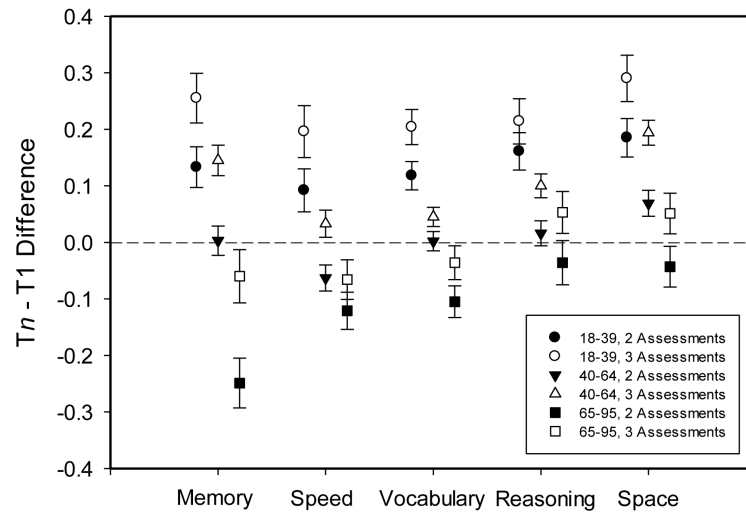


Figure 2. Estimated mean (and standard errors) composite score changes (i.e., $T_n - T_1$) for participants with two (filled symbols) and three assessments (open symbols) after control of the $T_1 - T_n$ interval and a measure of general cognitive ability (PC1).

Table 1
Demographic characteristics of participants with two or three assessments in three age groups

Number of Assessments	Age 18-39			Age 40-64			Age 65-95		
	Two	Three		Two	Three		Two	Three	
Sample Size	254	182		555	659		326	287	
Age	27.4 (6.7)	28.8 (7.0)		52.8 (6.4)	52.8 (6.7)		74.4 (6.6)	72.3 (5.7)	
Proportion Female	.66	.68		.68	.73		.60	.56	
Years Education	14.8 (2.4)	14.6 (2.2)		15.7 (2.6)	16.0 (2.6)		16.1 (2.8)	16.2 (3.0)	
Self-Rated Health	2.0 (0.8)	2.2 (0.9)		2.1 (0.9)	2.1 (0.9)		2.3 (0.9)	2.3 (0.9)	
Est. IQ	108.6 (12.4)	105.4 (15.4)		110.2 (14.6)	112.0 (15.1)		109.1 (13.1)	12.5 (13.0)	
T1-Tr Interval (Years)	3.5 (2.2)	5.7 (1.8)		3.6 (2.0)	5.9 (1.8)		2.9 (1.4)	5.4 (1.6)	
T1 Composite Scores									
Memory	.39 (.73)	.33 (.77)		.02 (.78)	.17 (.73)		-.46 (.76)	-.20 (.73)	
Speed	.62 (.70)	.62 (.73)		-.03 (.69)	.18 (.68)		-.78 (.70)	-.52 (.60)	
Vocabulary	-.31 (.88)	-.49 (.86)		.12 (.87)	.29 (.84)		.24 (.73)	.41 (.65)	
Reasoning	.46 (.75)	.25 (.89)		.07 (.81)	.14 (.80)		-.50 (.74)	-.26 (.73)	
Spatial Visualization	.45 (.90)	.30 (1.0)		-.03 (.76)	.08 (.79)		-.49 (.60)	-.33 (.68)	

Note: Health rating was on a scale from 1 for “excellent” to 5 for “poor”. Numbers in parentheses are standard deviations. T1 refers to the second occasion for participants with two assessments and to the third occasion for participants with three assessments.

Table 2

Results of analyses of variance on composite scores at first and final occasion with PC1 and total interval as covariates.

	Memory		Speed		Vocabulary		Reasoning		Spatial Vis.	
	F	η^2	F	η^2	F	η^2	F	η^2	F	η^2
Age	65.93*	.074	345.23*	.294	521.86*	.386	44.42*	.051	65.53*	.073
NumAssess	8.09*	.005	17.29*	.010	0.35	.000	0.09	.000	1.46	.001
Time	45.58*	.027	26.04*	.015	18.88*	.011	30.41*	.018	63.87*	.037
Age*NumAssess	0.06	.000	0.12	.000	2.79	.003	0.33	.000	1.06	.001
Age*Time	41.05*	.047	23.11*	.027	54.45*	.062	16.48*	.019	24.23*	.028
NumAssess*Time	17.80*	.011	7.37*	.004	8.06*	.005	5.24	.003	11.87*	.007
Age*NumAssess*Time	0.54	.001	0.45	.001	2.37	.003	1.71	.002	0.70	.001
All three sessions										
Age	44.41*	.087	189.39*	.287	259.82*	.368	48.68*	.106	46.61*	.102
NumAssess	2.05	.002	5.38	.006	1.03	.001	2.24	.003	5.82	.007
Time	32.22*	.033	20.15*	.021	13.72*	.015	23.56*	.028	16.07*	.010
Session	16.32*	.017	50.55*	.051	4.54	.005	2.32	.003	49.55*	.057
Age*NumAssess	0.82	.002	1.29	.003	1.29	.003	0.27	.001	2.69	.007
Age*Time	49.88*	.097	16.59*	.034	46.76*	.095	23.17*	.054	25.38*	.058
Age*Session	1.92	.004	5.40*	.011	113.21*	.202	0.30	.001	12.83*	.030
NumAssess*Time	11.78*	.012	3.20	.003	10.54*	.012	13.24*	.016	0.64	.001
NumAssess*Session	0.27	.000	4.64	.005	2.65	.003	1.09	.001	0.49	.001
Time*Session	8.35*	.009	0.36	.000	3.26	.004	1.34	.002	11.90*	.014
Age*NumAssess*Time	1.56	.003	0.32	.001	2.25	.005	3.25	.008	1.62	.004
Age*NumAssess*Session	0.82	.002	2.17	.005	0.30	.001	2.40	.006	1.75	.004
Age*Time*Session	4.33*	.009	1.37	.003	2.54	.006	1.10	.003	2.94	.007
NumAssess*Time*Session	2.41	.003	0.08	.000	1.72	.002	0.30	.000	2.06	.002
Age*NumAssess*Time*Session	0.49	.001	1.34	.003	0.69	.002	0.36	.001	0.28	.001

Note: NumAssess refers to the number of assessments (2 or 3). Time refers to the contrast between first (T1) and final (T2 or T3) occasion, and Session (or Sess) refers to the session within an occasion (i.e., first, second, or third).

* $p < .01$.