Intelligence 40 (2012) 352–361

Contents lists available at SciVerse ScienceDirect

Intelligence

Does the direction and magnitude of cognitive change depend on initial level of ability?

Timothy A. Salthouse

Department of Psychology, University of Virginia, Charlottesville, VA 22904–4400, United States

ARTICLE INFO

Article history:
Received 18 November 2011
Received in revised form 27 January 2012
Accepted 14 February 2012
Available online 7 March 2012

Keywords:
Aging
Longitudinal change
Ability-dependent change
Regression toward the mean

ABSTRACT

Longitudinal change in five cognitive abilities was investigated to determine if the direction or magnitude of change was related to the individual's ability level. Adults between 18 and 97 years of age performed three versions of 16 cognitive tests on two occasions separated by an average of 2.7 years. In order to control for influences associated with regression toward the mean, level of ability was determined from scores on the first version of the cognitive tests on the first occasion, and across-occasion change was examined on the second and third versions. Change in every cognitive ability was significantly more negative with increased age. However, there was little indication of ability-dependent change in any of the five cognitive abilities, either in differences between composite scores, or in estimates of latent change. Although there are reasons to expect cognitive change to be less negative at either high or low levels of ability, these data suggest that neither the direction nor magnitude of change is related to initial ability when influences of regression toward the mean are controlled.

1. Introduction

The primary questions in the current report were whether longitudinal change on cognitive tests operates in the same manner throughout the entire distribution of ability, and whether ability–change relations were similar at different ages. Less negative change might be expected among individuals with high levels of ability, either because they have a greater magnitude of some type of cognitive reserve (Stern, 2009), or because they have a lifestyle involving physical or cognitive activity conducive to the preservation of cognitive ability. It is also possible that individuals with low levels of ability have more negative change because they are in the early stages of neurodegenerative disease and the influences associated with the disease are superimposed on normal age-related influences.

However, there are at least two reasons why differential change might be expected in the opposite direction. First, individuals with low levels of ability may be close to a functional floor, in which case there may not be much room for negative change to be detected. And second, if the highest level of performance can only be achieved if all processes are functioning near optimum, individuals at higher ability levels might be more vulnerable to decline because disruption in any of several processes could lead to a reduction in performance, whereas lower levels of performance may be possible with many different combinations of processes.

Because average cognitive change has been found to differ as a function of age (e.g., Salthouse, 2010a,b), ability–change relations might also vary with age. For example, if different factors are contributing to change at different ages, and to the relations between initial ability and change, divergence of change according to initial ability might only occur at older ages, with more uniform change occurring at younger ages.

In a recent review of research on intelligence, Nisbett and colleagues stated: “Intelligence declines in old age, but do brighter individuals decline less or more? We do not have longitudinal data on the question …(Nisbett et al., 2012).”

E-mail address: salthouse@virginia.edu.

0160-2896/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2012.02.006

In fact, however, a number of studies have been published with longitudinal data, including four in which the phrase "Is age kinder to the initially more able" was included in the title. Much of this research was recently reviewed by Gow et al. (2012). They reported that the existing results were inconsistent, and suggested that small sample sizes, different ages at baseline and when change was assessed, and differences in the interval between assessments and in the nature of the cognitive tests could have contributed to the variability in results.

The Gow et al. (2012) review focused on studies in which the baseline measurement occurred in childhood or young adulthood because the authors suggested that studies with baselines in middle or late adulthood are difficult to interpret because cognitive declines may have already occurred. However, a somewhat different, but equally meaningful, question concerns the relation between an individual's current ability level and his or her subsequent cognitive decline, and for this question a short interval between baseline assessment and change is desirable. Furthermore, if the sample includes participants across a wide age range it is possible to determine if the ability–change relations vary as a function of age as one might expect if progressively more individuals are experiencing appreciable decline with increased age.

Several studies have examined the relation between level and change in ability when both the initial and change assessments have been in middle- or old-age. For example, Wilson, Beckett, Bennett, Albert, and Evans (1999) reported that older adults with lower levels on a composite measure of cognition at baseline had greater declines than older adults with higher initial levels. However, Reynolds, Finkel, Gatz, and Pedersen (2002) found different patterns of ability–change relations across different methods of analyzing change, and Wilson et al. (2002) found that the ability–change correlations ranged from .02 to .53 across composite scores representing different cognitive domains. Because the studies differed in the number and breadth of the cognitive domains evaluated as well as in the types of analytical procedures, it is difficult to determine reasons for the discrepancies in ability–change relations at this time.

It is useful to consider methods that can be used to examine the relation between initial ability and change. Some information about differential change can be obtained from comparisons of the ratio of the variance of scores on the first and second occasions, and from inspection of the correlation of the scores across the two occasions. The rationale is that if people differ in the direction and magnitude of longitudinal change, the variance on the second occasion should differ from that on the first occasion because the variance in change is superimposed on the initial variance, and correlations across the longitudinal interval (i.e., stability coefficients) should be less than 1.0 if the ordering of individuals differs on the two occasions. Although information of this type indicates whether the individual differences are greater in one occasion or the other, or whether the individual differences are stable across occasions, it is not directly relevant to the question of whether the direction and magnitude of change varies according to the individual's level of ability.

More direct information is available by comparing change in individuals selected from different portions of the ability distribution. However, it is not always recognized that the results from these types of analyses could be distorted by regression toward the mean. The problem is that if some of the high values on the first assessment are due to measurement error in the positive direction and some of the low values are due to measurement error in the negative direction, then the scores at the second assessment will tend to be less extreme than those on the first assessment for statistical, rather than substantive, reasons. In other words, individuals with high initial levels may exhibit some artifactual decline whereas individuals at low initial levels may exhibit artifactual gain. Unless influences of regression toward the mean are taken into consideration, therefore, the observed change will likely be composed of an unknown mixture of true change and of influences of regression toward the mean.

Several attempts have been made to either examine, or control, influences of regression toward the mean in analyses of cognitive change. For example, Baltes, Nesselroade, Schaie, and Labouvie (1972) used a time-reversed design to investigate the existence of regression toward the mean in longitudinal change in cognitive abilities. As noted by Alder, Adam, and Arenberg (1990), however, the time-reversed strategy merely indicates whether regression toward the mean is operating, and neither eliminates it, nor provides an estimate of its magnitude. Alder et al. (1990) used growth curve models with adjustments for measurement error to examine the relation between initial level and change in measures of vocabulary and visual memory. However, it is not clear how effective this method is in actually reducing influences of regression toward the mean, and the results were inconsistent because the estimated relations differed in direction for the two measures, with a positive ability–change relation for the visual memory measure but a negative relation for the vocabulary measure.

Because most regression occurs from the selection assessment to the next assessment, Nesselroade, Stigler, and Baltes (1980) proposed that effects associated with regression toward the mean could be minimized by using scores at the first assessment for selection, and then examining change

Table 1
Sample characteristics.

	Longitudinal		
	Only T1	Only T2 burst	T1 and T2 bursts
N	2322	1115	783
Age 18–39	765	223	148
Age 40–59	836	493	313
Age 60–79	580	351	268
Age 80–97	141	48	54
Age	49.2 (19.0)	53.0 (16.5)	54.9 (17.6)
Proportion female	.65	.66	.65
Self-rated health	2.2 (.9)	2.1 (.9)	2.3 (.9)
Years of education	15.6 (2.8)	15.7 (2.7)	15.7 (2.6)
Scaled scores			
Vocabulary	12.3 (3.2)	12.6 (3.1)	12.7 (2.9)
Digit symbol	11.1 (2.9)	11.5 (2.9)	11.5 (2.8)
Logical memory	11.5 (3.0)	12.0 (2.8)	11.9 (2.9)
Word recall	11.9 (3.3)	12.4 (3.2)	12.1 (3.5)
T11 − T21 (years)	NA	2.9 (1.4)	2.5 (.9)

Note: Values in parentheses are standard deviations. Health was a self-rating on a scale from 1 for excellent to 5 for poor. Scaled scores are adjusted for age and have means of 10 and standard deviations of 3 in the nationally representative normative samples (Wechsler, 1997a, 1997b).

beginning from the second assessment. Implementation of this staggered selection-assessment strategy therefore requires a minimum of three assessments, one for selection and two to assess change. In a traditional longitudinal study with only a single assessment at each occasion, application of this strategy will result in loss of change information originating from the first occasion, which is unfortunate because early change could be different than later change.

An alternative implementation of the staggered selection-assessment procedure can be used with a measurement burst design, such as that incorporated in the Virginia Cognitive Aging Project (VCAP; Salthouse, 2007, 2010a). VCAP is an ongoing cross-sectional and longitudinal study of cognitive aging that began in 2001. A unique feature of the project is a measurement burst design in which three assessments on parallel versions of 16 tests are administered at each longitudinal occasion. Because multiple assessments are available at each occasion, individuals at different levels of ability can be selected on the basis of the scores on the first assessment at the first occasion, and then change examined from the first to the second occasion on subsequent assessments. That is, assume that the three assessments at Time 1 are designated T11, T12, and T13, and those at Time 2 are designated T21, T22, and T23. The staggered selection-assessment procedure can therefore be implemented in a measurement burst design by determining initial ability on the basis of T11 scores, and then examining change from T12 to T22 and from T13 to T23.

2. Methods

2.1. Participants

All participants in VCAP who returned for a second occasion were administered the three-session measurement burst. However, on the first occasion some of the participants were administered the VCAP battery only on the first session and were administered different cognitive tests on the second and third sessions, and therefore these participants had only one measurement burst. Characteristics of the longitudinal participants with one and two measurement bursts, and of the participants with only data at the first (T1) occasion, are reported in Table 1. The self-identified ethnicity of the participants was primarily White (81%) with about 10% reporting themselves to be Black, and the remainder split among different groups including mixed ethnicity. The T1 and T2 testing was conducted between 2001 and 2011, but there were little or no effects of test year on performance, and therefore the data were aggregated across all years.

Inspection of Table 1 reveals that most participants rated their health in the very good range, and that the average years of education was over 15. The age-adjusted scaled scores from four standardized tests indicate that participants in the sample performed between .5 and 1 standard deviations above the average in the nationally representative normative samples. The participants in this sample can therefore be inferred to be relatively high functioning, both in terms of self-rated health and level of cognitive ability. The longitudinal participants were slightly older than individuals completing only one occasion, likely because of greater mobility in the younger groups (Salthouse, 2010a). The participants completing two occasions also had higher scores on two memory measures involving recall of unrelated words and meaningful stories than participants with only one occasion.

Table 2
Comparisons of results with composite scores at T11 and T21 with all longitudinal participants.

	T21 − T11 difference	Variance ratio	Correlations	
	Mean (SD)	T21/T11	1 week	2.7 years
Memory				
Age 18 − 39	.15* (.47)	1.16	.84	.80
Age 40 − 59	.02 (.50)	.98	.86	.76
Age 60 − 79	−.11* (.49)	.99	.87	.74
Age 80 − 97	−.17 (.56)	.78	.84	.63
Speed				
Age 18 − 39	.05 (.48)	1.29	.92	.80
Age 40 − 59	−.04 (.46)	1.10	.88	.79
Age 60 − 79	−.11* (.41)	.97	.88	.81
Age 80 − 97	−.21* (.42)	1.10	.87	.76
Vocabulary				
Age 18 − 39	.10 (.28)	.97	.96	.94
Age 40 − 59	.02 (.28)	.94	.95	.94
Age 60 − 79	−.06 (.32)	1.03	.95	.88
Age 80 − 97	−.12 (.44)	.95	.92	.80
Reasoning				
Age 18 − 39	.15* (.40)	.96	.89	.87
Age 40 − 59	.06* (.39)	.97	.92	.88
Age 60 − 79	.03 (.40)	1.01	.85	.84
Age 80 − 97	.01 (.44)	1.07	.91	.78
Space				
Age 18 − 39	.20* (.40)	1.02	.91	.91
Age 40 − 59	.10* (.43)	1.04	.91	.85
Age 60 − 79	.01 (.39)	1.03	.83	.84
Age 80 − 97	−.06 (.34)	.70	.67	.77

Note: *p<.01. Correlations with the 1-week interval were based on data from Study 3 in Salthouse and Tucker-Drob (2008).

2.2. Tests

A total of 16 cognitive tests, representing five cognitive abilities, were administered in the same order to all participants. Episodic memory was represented by word recall, paired associates, and story memory tests. Perceptual speed ability was represented by a digit symbol substitution test, and pattern comparison and letter comparison tests. Vocabulary was represented by a

Table 3
Across-version (session) correlations of composite scores and of latent constructs.

	Composites				Latent Construct		
	11–11	11–12	11–13	12–13	11–12	11–13	12–13
Memory	.90	.83	.79	.84	.95	.95	.98
Speed	.94	.91	.90	.93	.99	.99	.99
Vocabulary	.95	.87	.85	.83	.97	.97	.96
Reasoning	.91	.87	.86	.85	.99	.98	.98
Space	.90	.84	.79	.78	.99	.99	.99

Note: The columns indicate which sessions/versions were correlated. For example, the 11–12 columns contain correlations between scores on the first version at the first occasion (11) and scores on the second version at the first occasion (12). Correlations with the same version (11–11) are based on data from Study 3 in Salthouse and Tucker-Drob (2008).

provide-the-definition test, a picture naming test, and multiple-choice synonym and antonym tests. Reasoning was represented by a matrix reasoning test, a letter sets test, and a series completion test. Spatial visualization ability was represented by a spatial relations test, a paper folding test, and a form boards test. Details of the tests, including reliabilities and results of confirmatory factor analyses supporting the hypothesized ability structure, are reported in other publications (Salthouse, 2007, 2010a,b,c; Salthouse, Pink, & Tucker-Drob, 2008).

Three different versions of each of the 16 tests were created with different items, but an identical format. Because the focus in VCAP is on individual differences, and varying the order of the tests across participants would introduce a confounding between characteristics of the individual and how he or she was treated in the testing situation, the versions were administered in the same fixed order to all participants. However, it is possible that the test versions could differ in mean performance, which would complicate direct comparisons across versions. The three versions were therefore administered in a counterbalanced order to a separate sample of 90 adults between 20 and 79 years of age to determine the average level of performance in each version without a confounding of version and presentation order (Salthouse, 2007). Regression equations in this sample were used to predict performance in the original test version from performance in the second or third versions, and the intercepts and slopes of these equations were then used to adjust the scores on the second and third versions of every participant in the current study to remove any order-independent version differences in the means (cf. Salthouse, 2007). Finally, in order to facilitate comparisons across tests, all test scores were converted into the same scale based on the distribution of first assessment scores by subtracting the mean and dividing the difference by the standard deviation.

2.3. Analytical procedures

Two analytical approaches based on the staggered selection-assessment procedure were used to investigate whether longitudinal change differed according to initial ability level. The first method consisted of dividing the participants into quartiles on the basis of T11 scores, and then using quartile and time (i.e., T1 or T2) as factors in a mixed (between-subjects for quartile, within-subjects for time) ANOVA comparing scores on the session 2 (i.e., T12 and T22) or session 3 (i.e., T13 and T23) assessments. Of particular interest in these analyses is the interaction of quartile by time because it provides a test of whether longitudinal change varies according to initial ability level.

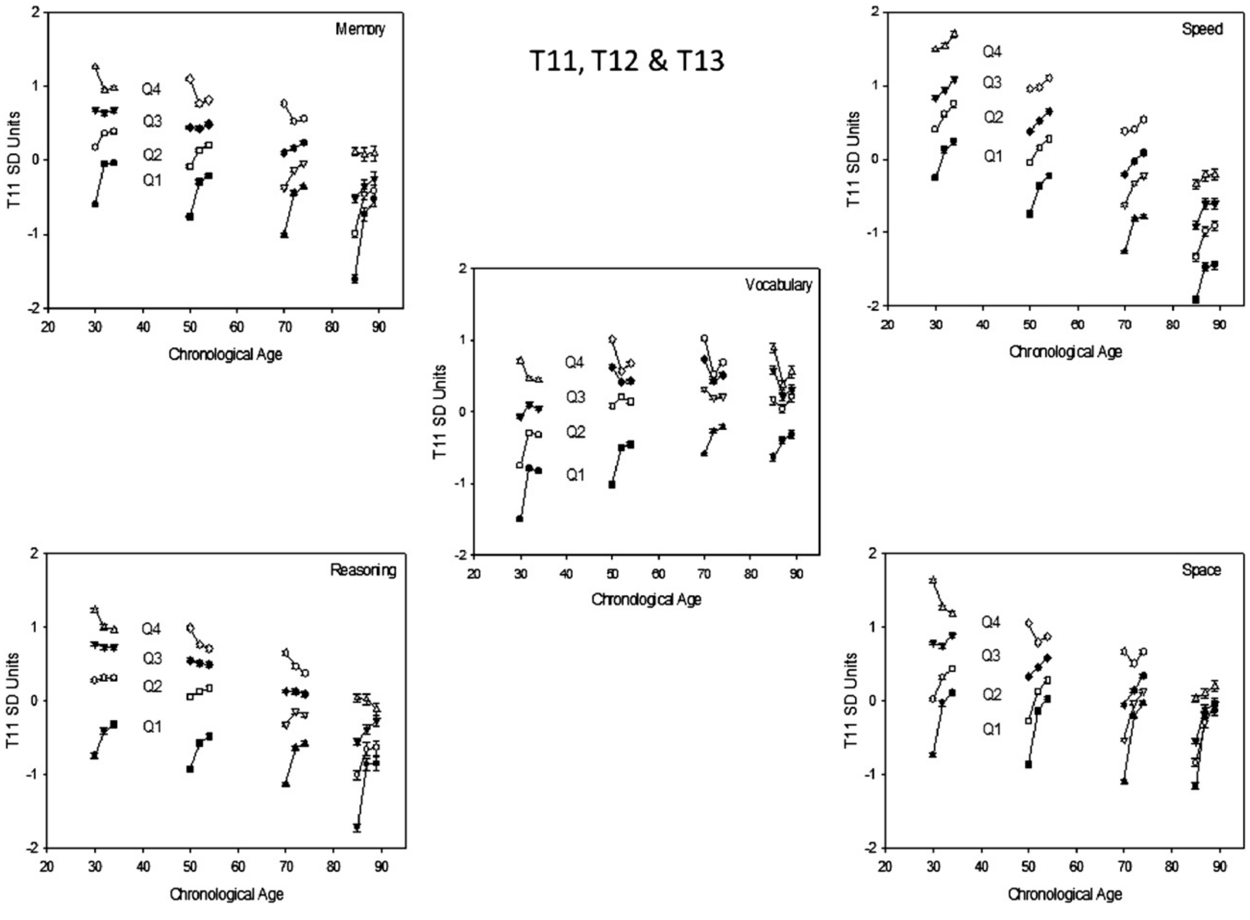

Fig. 1. Means and standard errors of composite scores on the first (T11), second (T12), and third (T13) sessions of the first occasion for participants in four quartiles based on T11 scores in four age groups. Although portrayed with intervals of 2 years to improve visibility, the actual intervals between sessions averaged about 7 days.

The second analytical approach consisted of using latent change models (cf. Ferrer & McArdle, 2010) based on the three (or four, for vocabulary ability) indicator variables of each ability at each occasion on either session 2 or session 3. Moreover, instead of categorizing the T11 selection variable into quartiles, the T11 score was treated as a continuous variable postulated to influence the level and change parameters in the model. (Both linear and quadratic relations of the T11 variables were examined, but no quadratic trends were significant, and therefore they were dropped, and the analyses repeated with only the linear relation.) Positive coefficients relating T11 score to change would indicate that the change is more positive (or less negative) at higher T11 levels, and negative coefficients would indicate that the change is more negative (or less positive) at higher T11 levels.

3. Results

As noted earlier, scores on the second and third versions of each test were adjusted with regression equations to minimize across-version differences in mean performance, and all scores converted to z-scores based on the mean and standard deviation of the T11 scores. Composite scores were then created at each session by averaging z-scores for the tests representing each ability.

The initial analyses were conducted on the T11 and T21 composite scores from all 1898 longitudinal participants. The first column of Table 2 contains means and standard deviations of the T21–T11 differences in the four age groups. Not all of the mean differences were significantly different from zero, but for every composite score the mean difference was less positive (or more negative) at older ages. In contrast, no systematic age trend was evident for the standard deviations of the T21–T11 differences between composite scores.

The second column of Table 2 contains ratios of the variance of the composite scores in the second occasion to the variance in the first occasion. Ratios above 1.0 would signify a fan effect on change in which the variance at the second occasions was larger than that on the first occasion. It can be seen that all of the values were close to 1, and that the highest ratios, indicating the greatest time-related increases in variance, were in the youngest group with memory and speed abilities.

The third and fourth columns in Table 2 contain correlations between the composite scores across the two occasions. The entries in the third column are based on data from 227 adults in Study 3 of Salthouse and Tucker-Drob (2008) who performed identical versions of the tests with an average interval of about 1 week. These values provide a baseline to assist in the interpretation of correlations of the same test versions across the longitudinal interval because even with very short retest intervals the correlations are not 1.0. The 1 week and 2.5 year correlations were very similar for space ability, but for the other abilities the correlations were smaller in the longitudinal data than in the 1 week retest data, indicating a shift in the ordering of individuals during the longitudinal interval. The correlations were nearly constant at all ages for speed ability, but the correlations over the 2.7 year interval were smaller at older ages for memory, vocabulary, and reasoning, indicating lower stability of the individual differences with increased age.

In order to compare change across test versions (and sessions), evidence is needed that the different versions are assessing the same construct. Correlations of composite scores and of latent constructs for each ability were therefore computed across versions (sessions). For comparison purposes, results with the 1-week retest correlations involving the same test versions from Salthouse and Tucker-Drob (2008) are also reported. The results, presented in Table 3, indicate that the correlations between different versions were only slightly smaller than the correlations for identical versions, and that all of the correlations for the latent constructs were very close to 1.0. It therefore seems reasonable to conclude that the different test versions assess very similar, if not identical, ability constructs.

Composite scores for each ability at T11, T12, and T13 are portrayed in Fig. 1 for participants divided according to T11 quartiles in the four age groups. Only data from the 783 participants with two 3-session measurement bursts were used in these across-session comparisons. Because the standard error bars are quite small, the estimates can be inferred to be fairly precise. Many of the patterns in Fig. 1 are consistent with regression toward the mean from the first to the second assessment as there were decreases from session 1 to session 2 at high ability levels (fourth quartile, Q4) and increases at low ability levels (first quartile, Q1). However, there was little evidence of a regression effect between sessions 2 and 3. Speed ability is an exception to the preceding generalizations as participants at all quartiles exhibited moderate to large

Table 4
F values and partial eta squared (η^2) effect sizes for interactions of quartile at T1 and session at the first occasion on composite scores.

	Sessions 11 and 12		Sessions 12 and 13	
	F	η^2	F	η^2
Memory				
Age 18–39	147.34*	.48	.91	.00
Age 40–59	187.64*	.43	.74	.00
Age 60–79	146.20*	.44	.86	.01
Age 80–97	27.29*	.49	1.22	.06
Speed				
Age 18–39	18.55*	.10	.64	.00
Age 40–59	39.07*	.13	.18	.00
Age 60–79	55.79*	.21	4.55*	.02
Age 80–97	12.36*	.22	.56	.01
Reasoning				
Age 18–39	47.43*	.25	2.48	.02
Age 40–59	90.87*	.29	6.68*	.03
Age 60–79	90.05*	.35	3.79*	.02
Age 80–97	26.57*	.51	1.51	.06
Space				
Age 18–39	239.22*	.61	11.15*	.08
Age 40–59	372.19*	.61	2.25	.01
Age 60–79	253.53*	.58	.35	.00
Age 80–97	30.30*	.54	1.50	.06
Vocabulary				
Age 18–39	270.65*	.64	.44	.00
Age 40–59	508.59*	.68	9.63*	.04
Age 60–79	263.21*	.57	4.73*	.02
Age 80–97	43.13*	.52	.69	.02

* $p < .01$

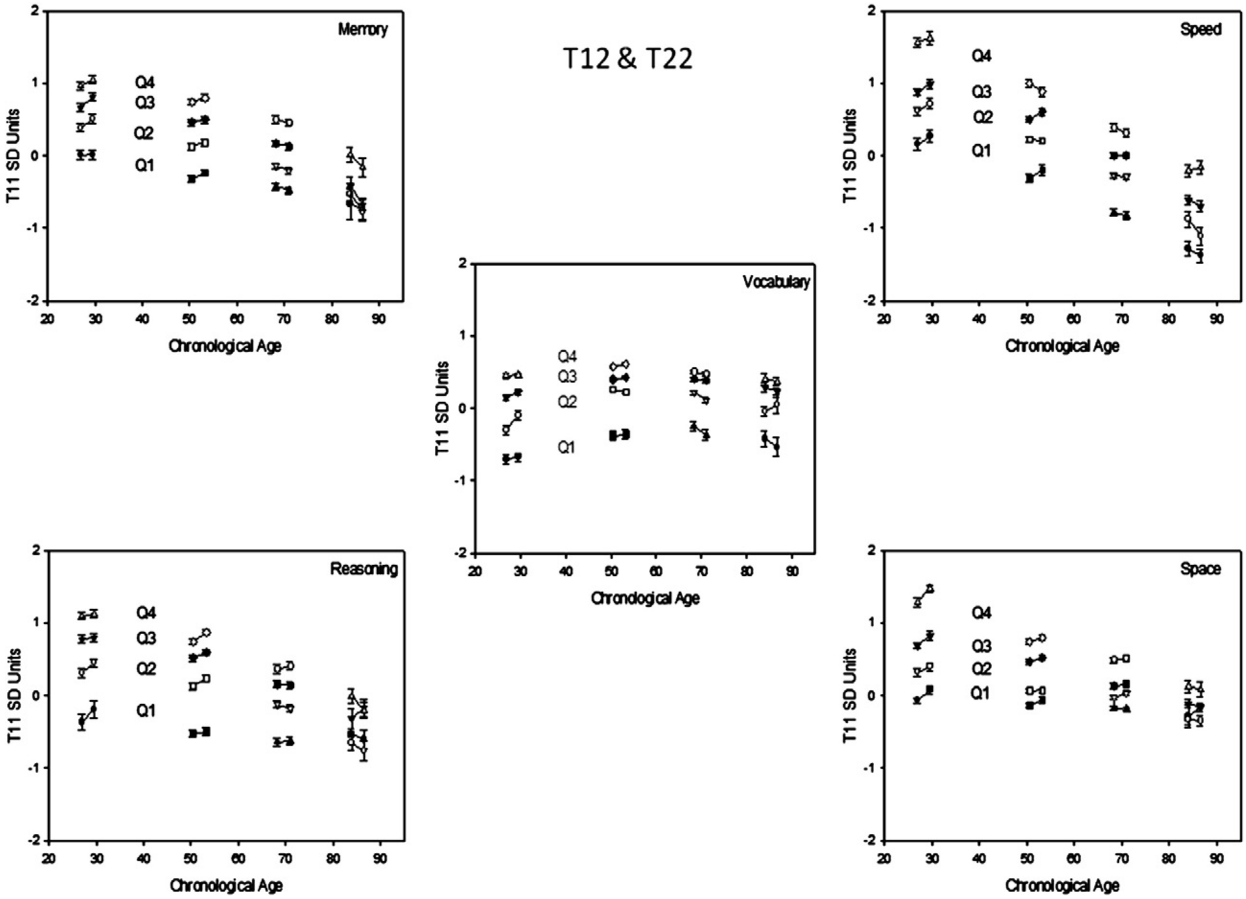

Fig. 2. Means and standard errors of composite scores on the second session of the first (T12) and second (T22) occasions for participants in four quartiles based on T11 scores in four age groups.

gains from session 1 to session 2 as well as some additional gains from session 2 to session 3.

Results of the interactions of ability quartile and session (1 vs. 2, and 2 vs. 3) in analyses of variance for each ability and each age group are reported in Table 4. The results confirm the impressions of Fig. 1 as all of the interactions were large and significant for the contrast between sessions 1 and 2, but were small and generally not significant for the contrast between sessions 2 and 3.

Fig. 2 portrays the T12 and T22 composite scores, reflecting across-occasion change on the second assessment, by T11 quartiles, representing ability selection at the first assessment, in the four age groups. (Although not illustrated to conserve space, the patterns with the T13 and T23 scores were very similar to those with the T12 and T22 scores.) The finding that most of the lines representing longitudinal change were nearly parallel indicates that there is little evidence of a systematic pattern of differential change by ability quartile. Moreover, the small standard errors of the estimates suggest that the lack of differential trends was not simply attributable to measurement imprecision.

The data summarized in Fig. 2 were analyzed with Quartile × Time (T1, T2) analyses of variance separately on the session 2 and session 3 composite scores in each age group and each cognitive ability. The F-ratios and estimated effect sizes for the quartile × time interactions for each cognitive ability and age group are presented in Table 5. It can be seen that of the 40 possible interactions (i.e., changes in 2 sessions in each of 5 abilities and 4 age groups), only the interactions for the speed composite score with session 2 change in the age 40–59 group, and for the vocabulary composite score with session 3 change in the 18–39 group were significant. These data are consistent with the nearly parallel lines connecting the T1 and T2 occasions in Fig. 2.

The next set of analyses consisted of latent change models in which the three or four (for vocabulary) measures of each ability served as indicators of the latent constructs on each session. Although not all longitudinal participants had measurement bursts at both occasions (cf. Table 1), latent change models were evaluated using the full information maximum likelihood (FIML) algorithm that incorporates all available information in the analyses (Arbuckle, 2007), including that from participants with data on only one occasion. The FIML procedure is based on the assumption that the data are missing at random, which seems plausible in this situation because most of the missing data was attributable to the particular combination of tests administered to the participant.

Most fits of the latent change models were in the moderate range, with comparative fit index (CFI) values above .90, and root mean square error of approximation (RMSEA) values below .10. However, the fits were quite poor (e.g.,

Table 5
F values and partial eta squared (η^2) effect sizes for interactions of quartile at T1 and time (T1 vs. T2) on composite scores.

	Session 2		Session 3	
	F	η^2	F	η^2
Memory				
Age 18–39	1.33	.03	.08	.00
Age 40–59	.28	.00	1.16	.01
Age 60–79	.74	.00	.20	.00
Age 80–97	.41	.04	.98	.12
Speed				
Age 18–39	.09	.00	.16	.00
Age 40–59	5.23*	.05	1.05	.01
Age 60–79	.80	.01	1.16	.01
Age 80–97	1.60	.09	.35	.02
Reasoning				
Age 18–39	2.06	.05	2.53	.07
Age 40–59	.80	.01	.11	.00
Age 60–79	1.02	.02	2.36	.03
Age 80–97	1.68	.15	.81	.07
Space				
Age 18–39	.70	.02	.18	.01
Age 40–59	.67	.01	.55	.01
Age 60–79	.75	.01	.67	.01
Age 80–97	.54	.05	.52	.04
Vocabulary				
Age 18–39	3.70	.08	6.61*	.14
Age 40–59	2.11	.02	1.23	.01
Age 60–79	1.78	.02	.20	.00
Age 80–97	1.21	.08	.43	.03

* $p<.01$.

CFI<.50) for memory ability in the oldest (age 80 to 97) age group, and therefore no estimates are presented in this age–ability combination. The latent change estimates for change in sessions 2 and 3 are portrayed in Fig. 3 for the other age–ability combinations, where it can be seen that the patterns of change were very similar in the session 2 and session 3 assessments. Furthermore, with the exception of speed ability in which all change estimates were negative, the estimated changes were positive at younger ages and more negative at older ages.

Relations of age to the latent change estimates were also examined across all ages in the complete sample. The standardized coefficients relating age to the estimates for the session 2 and session 3 changes, were, respectively, −.36 and −.43 for memory ability, −.12 and −.15 for speed ability, −.49 and −.46 for vocabulary ability, −.23 and −.26 for reasoning ability, and −.19 and −.28 for space ability. All relations were significantly ($p<.01$) less than zero, with generally similar values in the two sessions in each ability.

Fig. 4 illustrates relations of initial ability at T11 on the estimates of change from T12 to T22 and from T13 to T23. As with the estimates of mean change in Fig. 3, the patterns, in this case of the relations of T11 ability to change, were similar with the changes in both sessions 2 and 3. All of the relations were rather small, with a few slightly negative, indicating less positive (or more negative) change among individuals at higher ability levels.

In order to explore the generalizability of the weak ability–change relations, two more general measures of cognitive ability were also examined. One was the score on the 35-item version of the North American Adult Reading Test (NAART, Uttl, 2002), that has been considered an indicator of the individual's peak ability early in life (e.g., Deary, Whalley, & Crawford, 2004). The other general measure of cognitive ability was the individual's score on the first principal component (1st PC) based on the T11 scores because the 1st PC is sometimes used as an estimate of general cognitive ability (Jensen, 1998).

Age, NAART or 1st PC, and the cross-product terms of age and either NAART or the 1st PC representing the interactions, were included as simultaneous predictors of latent change in these analyses to partial influences of age when considering effects of the predictors, and to determine if the influences of NAART or 1st PC varied with age. None of the interactions involving age was significant, and only two out of 10 (i.e., changes on sessions 2 and 3 for each of 5 abilities) relations were significant ($p<.01$). The significant effects were in the direction of less negative memory change on session 2 among individuals with higher NAART scores, and less negative space change on session 2 among individuals with higher 1st PC scores. With these two exceptions, the pattern resembled that with the narrower construct-specific measures of ability in that there was little evidence of ability-dependent cognitive change.

3. Discussion

Fig. 1 dramatically illustrates the regression-toward-the-mean phenomenon as the change from the first (T11) to the second (T12) assessments was negative among individuals in the top quartiles of initial ability and was positive among individuals from the bottom quartile. These results are intriguing because regression toward the mean is often assumed to be small when the measurements are reliable, and yet the high alternate-form correlations in Table 3 indicate that the composite scores in this study were quite reliable. In order to reach precise quantitative estimates of ability–change relations, therefore, the phenomenon of regression-toward-the-mean needs to be considered in any analysis involving selection of individuals from extremes of the distribution.

Many of the measures of cognitive change in this study were significantly different from zero, including the T2–T1 differences in the composite scores in Table 2, and the latent change estimates in Fig. 3. However, in none of the measures was there much evidence of differential change as the variance on the second occasion was similar in magnitude to that on the first occasion, and the stability coefficients over the 2.7 year longitudinal interval were only slightly smaller than the correlations over a 1 week interval. Most importantly, two different types of analyses examining relations between initial ability and change revealed little indication of greater change among individuals at either high or low levels of initial ability. Furthermore, despite a systematic shift in the direction and magnitude of change with increased age, the patterns of small to non-existent ability-dependent change were very similar at all ages, and with measures from five distinct cognitive abilities.

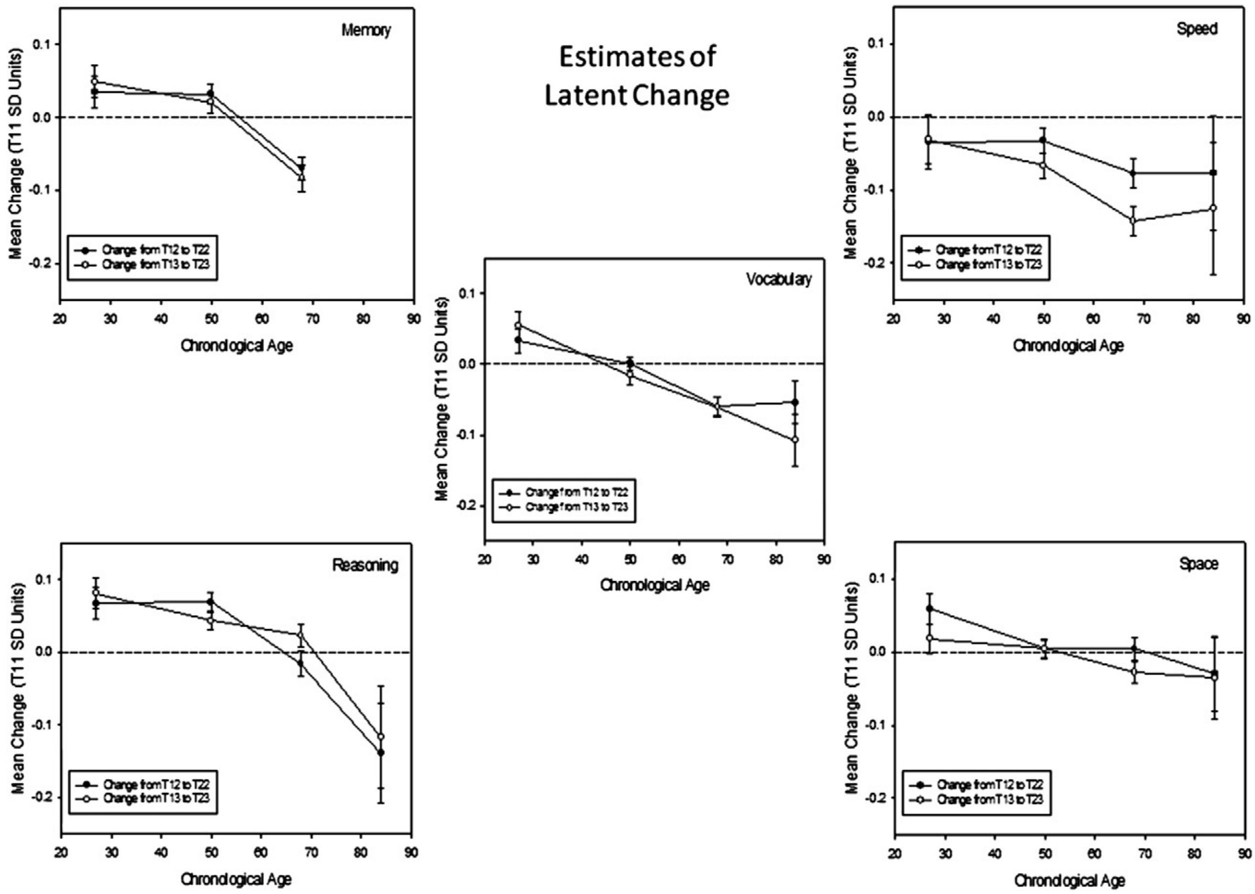

Fig. 3. Means and standard errors of latent change estimates on session 2 and session 3 in four age groups. No values are presented for memory ability in the oldest age group because the models had very poor fit to the data.

The weak support for differential longitudinal change associated with the initial level of ability might be considered surprising because the participants in the current study spanned a considerable range of ability, and the wide distribution of ages was associated with quite different average values of change. Statistical power is always a concern when reporting null results, but several considerations suggest that this is not a serious problem in the current study. First, there was little indication in Figs. 2 and 4 of systematic patterns in which change is related to initial ability. Second, the sample sizes were moderately large, and the small standard errors in the figures indicate that most of the estimates were quite precise. And third, the results were broadly consistent across different analytical methods, and with two separate measures of change based on the across-occasion changes on the second and third sessions, in five cognitive abilities, and in four different age groups.

How can the inconsistencies in prior research be explained? One possibility is that some of the earlier studies contained large numbers of adults over the age of 65, and thus the samples could have included individuals with age-related pathologies who may have been experiencing pronounced decline. Another possibility is that in some studies the relations were not very precise because cognitive functioning was assessed with single variables that are subject to measurement error. There has also been considerable variation in the nature of the analytical procedures, and both Reynolds, Gatz, and Pedersen (2002) and Gow et al. (2008) found different patterns of ability–change relations with different methods of analyzing change. Finally, some of the variability in results could be attributable to differences in the magnitude of between-person variance in change. Individual differences in change are often quite small, particularly when measurement error is minimized with latent constructs, and this could have limited detection of differential change (Gow et al., 2008).

It is also important to recognize that none of the prior studies incorporated systematic control of regression-toward-the-mean effects, and the current results indicate that this phenomenon can distort estimates of ability-dependent change even with reliable composite variables. Effects of regression toward the mean could also be operating in many latent change and growth curve analyses in which relations of the intercept and slope are examined. That is, negative relations between the intercept and slope (e.g., Ferrer, Salthouse, McArdle, Stewart, & Schwartz, 2005; Infurna, Gerstorf, Ryan, & Smith, 2011; Reynolds et al., 2002, reflecting more positive change at lower ability levels, could be completely attributable to regression toward the mean,

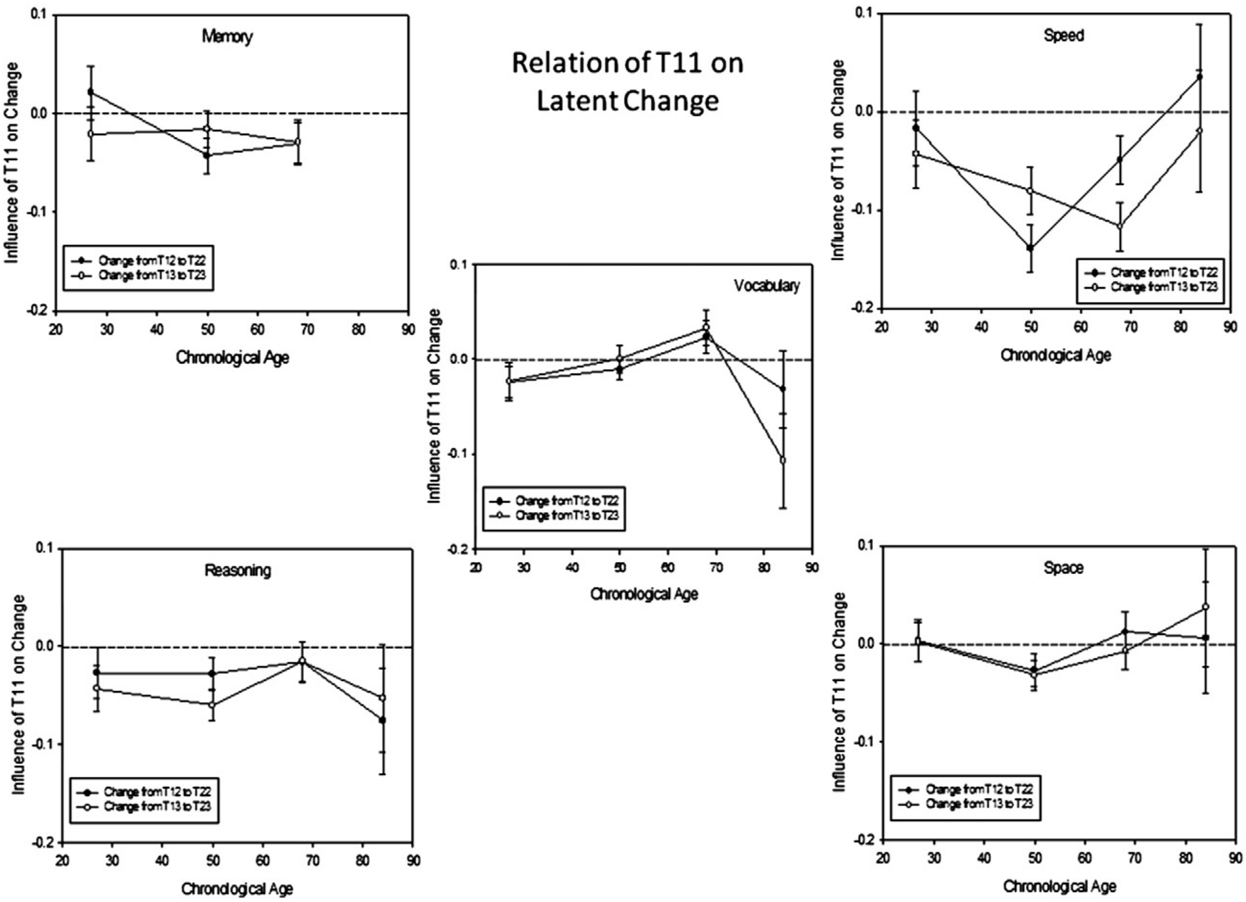

Fig. 4. Means and standard errors of regression coefficients relating T11 score to latent change on sessions 2 and 3 in four age groups. No values are presented for memory ability in the oldest age group because the models had very poor fit to the data.

and positive relations (e.g., Bielak, Gerstorf, Kiely, Anstey, & Luszcz, 2011; Sliwinski, Hofer, & Hall, 2003) could be underestimates of the true effects because they may be attenuated by regression effects. Methods similar to those employed here are therefore desirable to ensure that any differential change that might be observed is not an artifact of, or attenuated by, regression toward the mean.

To summarize, the major finding in this study is that after controlling for influences of regression toward the mean, little or no relations between level of ability and either the direction or magnitude of cognitive change are evident. Moreover, this is true for five different cognitive abilities, and despite more negative change at older ages, at four different periods in adulthood. Together with evidence of weak to non-existent relations between age and variance in cognitive change (Salthouse, 2010b, 2011, in press), these results are consistent with the possibility that the influences contributing to the negative relations between age and change are both larger, and somewhat different, than the factors responsible for individual differences in change at any age. An important question for future research is to determine the mechanisms responsible for the systematic relations of age to cognitive change in the absence of increases in variance, or differential change across different regions of the ability distribution.

Acknowledgments

This research 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. There are no conflicts of interest.

I would like to thank John R. Nesselroade for valuable comments on an earlier draft of this manuscript.

References

Alder, A. G., Adam, J., & Arenberg, D. (1990). Individual-differences assessment of the relationship between change in and initial level of adult cognitive functioning. *Psychology and Aging*, *5*, 560–568.

Arbuckle, J. L. (2007). *AMOS (version 7), [computer program].* Chicago: SPSS.

Baltes, P. B., Nesselroade, J. R., Schaie, K. W., & Labouvie, E. W. (1972). On the dilemma of regression effects in examining ability-level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology*, *6*, 78–84.

Bielak, A. A. M., Gerstorf, D., Kiely, K. M., Anstey, K. J., & Luszcz, M. (2011). Depressive symptoms predict decline in perceptual speed in older adulthood. *Psychology and Aging*, *26*, 576–583.

Deary, I. J., Whalley, L. J., & Crawford, J. R. (2004). An "instantaneous" estimate of a lifetime's cognitive change. *Intelligence*, *32*, 113–119.

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, *19*, 149–154.

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.

Gow, A. J., Johnson, W., Mishra, G., HALCyon Study Team, Richards, M., Kuh, D., & Deary, I. J. (2012). Is age kinder to the initially more able?: Yes, and no. *Intelligence*, *40*, 49–59.

Gow, A. J., Johnson, W., Pattie, A., Whiteman, M. C., Starr, J., & Deary, I. J. (2008). Mental ability in childhood and cognitive aging. *Gerontology*, *54*, 177–186.

Infurna, F. J., Gerstorf, D., Ryan, L. H., & Smith, J. (2011). Dynamic links between memory and functional limitations in old age: Longitudinal evidence for age-based structural dynamics from the AHEAD study. *Psychology and Aging*, *26*, 546–558.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, *88*, 622–637.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2002). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159.

Reynolds, C. A., Finkel, D., Gatz, M., & Pedersen, N. L. (2002). Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth models. *Experimental Aging Research*, *28*, 407–433.

Reynolds, C. A., Gatz, M., & Pedersen, N. L. (2002). Individual variation for cognitive decline: Quantitative methods for describing patterns of change. *Psychology and Aging*, *17*, 271–287.

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

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

Salthouse, T. A. (2010b). The paradox of cognitive change. *Journal of Clinical and Experimental Neuropsychology*, *32*, 622–629.

Salthouse, T. A. (2010c). Does the meaning of neurocognitive change change with age? *Neuropsychology*, *24*, 273–278.

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

Salthouse, T.A. (in press). Are individual differences in rates of aging greater at older ages? *Neurobiology of Aging*.

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*, 464–486.

Salthouse, T. A., & Tucker-Drob, E. M. (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*, *22*, 800–811.

Sliwinski, M. J., Hofer, S. M., & Hall, C. (2003). Correlated and coupled cognitive change in older adults with and without preclinical dementia. *Psychology and Aging*, *18*, 672–683.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*, 2015–2028.

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1123–1137.

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., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., et al. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, *17*, 179–193.

Wilson, R. S., Beckett, L. A., Bennett, D. A., Albert, M. S., & Evans, D. A. (1999). Change in cognitive function in older persons from a community population. *Archives of Neurology*, *56*, 1274–1279.