



Effects of age and ability on components of cognitive change



Timothy A. Salthouse*

Department of Psychology, University of Virginia, 102 Gilmer Hall, Charlottesville, VA 22904-4400, United States

ARTICLE INFO

Article history:

Received 1 May 2013

Received in revised form 3 July 2013

Accepted 4 July 2013

Available online xxxx

Keywords:

Aging

Longitudinal

Change

Retest

Cognitive abilities

ABSTRACT

Prior experience with a cognitive task is often associated with higher performance on a second assessment, and these experience effects can complicate the interpretation of cognitive change. The current study was designed to investigate experience effects by obtaining measures of cognitive performance separated by days and by years. The analyses were based on data from 2017 adults with two longitudinal occasions, of whom 948 had also completed a third occasion, with each occasion consisting of three parallel versions of the tests on separate sessions. Change across short intervals was typically positive, and greater among older adults and adults with low levels of cognitive ability, whereas change over intervals of approximately three years was often negative, particularly at older ages. In contrast to the expectation that change over short intervals might be informative about change over longer intervals, relations between short-term change and long-term change were negative, as the individuals who gained the most with assessments separated by days tended to experience the greatest losses across assessments separated by years.

© 2013 Published by Elsevier Inc.

1. Introduction

Traditional longitudinal comparisons involve at least two measurement occasions, with change assessed by contrasts among the scores from single assessments at each occasion. Longitudinal change is often interpreted as reflecting processes occurring over the interval between occasions, such as those related to development or maturation. However, because in longitudinal studies a second occasion is necessarily preceded by an initial occasion, some of the change could be attributable to prior experience with the tests. Developmental effects and test experience effects are difficult to separate in traditional longitudinal designs, but the two components of change can be distinguished if the research design involves multiple assessments at each occasion, as with dual-baseline procedures (e.g., Beglinger et al., 2005; McCaffrey & Westervelt, 1995; Van Gorp, Lamb, & Schmitt, 1993), or measurement burst designs (e.g., Nesselroade, 1991; Salthouse & Nesselroade, 2010). Very few studies have been reported with either type of design, but both could be informative in distinguishing components of change. Dual-baseline procedures

differ from conventional longitudinal designs by having two or more assessments at the initial occasion, and measurement burst designs differ by having a burst of multiple assessments at each occasion instead of a single assessment.

The top panel of Fig. 1 illustrates a traditional longitudinal comparison with only a single assessment at each occasion, and the bottom panel portrays a measurement burst design with three assessments (administered on separate sessions) at each occasion. Assessments in a measurement burst design can be designated by two numbers, with the first number referring to the occasion and the second referring to the session within an occasion. For example, 11 refers to the first session in the first occasion, 13 refers to the third session in the first occasion, and 22 refers to the second session in the second occasion. Note that in a traditional longitudinal comparison, change corresponds to the contrast between 11 and 21 because there is only a single assessment at each occasion. However, when three assessments are available at each occasion the longitudinal change (i.e., from 11 to 21) can be partitioned into components corresponding to change from 11 to 12, 12 to 13, and 13 to 21. The first two contrasts are within-occasion changes, whereas the third contrast represents the change from the last assessment in the first occasion to the first assessment in the second occasion.

* Tel.: +1 434 982 6323.

E-mail address: Salthouse@virginia.edu.

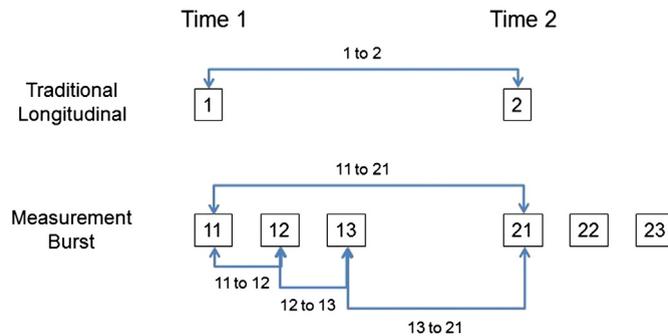


Fig. 1. Schematic illustration of possible measures of change in a traditional longitudinal study (top panel), and in a three-assessment measurement burst design (bottom panel).

The availability of multiple measures of change allows three important questions to be asked. First, can the contributions of different factors to cognitive change be assessed by contrasts of change across short (i.e., about one week) and longer (i.e., about 3 years) intervals? Second, do the measures of change across different intervals differ in their patterns of relations with individual difference characteristics, as might be expected if they reflect distinct aspects of change? And third, what is the relation between change over short and long intervals?

Four factors differing in their probable degree of generalizability can be postulated to contribute to change in measures of cognitive functioning. One factor is memory for specific items in the tests, which will likely have its greatest effect when identical test versions are used in each assessment. A second factor that could be involved in change is the development of test-specific skills or strategies, which could affect performance even when the tests in successive assessments involve different items. A third factor that could be contributing to change is an increase or decrease in the relevant construct or ability, in which case effects would be expected on different tests of the same ability. Finally, some change may be attributable to shifts in construct-irrelevant factors such as anxiety or unfamiliarity with testing, which might have effects on any type of cognitive test, and not merely those evaluating the same cognitive ability.

The contribution of memory for specific items can be evaluated with a comparison of change involving identical or different test versions. However, information about the contributions of the other factors might be obtained by comparing change across different intervals. For example, change over short intervals with different test versions at each assessment might primarily reflect the acquisition of test-specific skills or strategies and/or reduction in anxiety and unfamiliarity, whereas change over longer intervals may be more likely to reflect change in the relevant ability (cf. Salthouse, 2009; Salthouse & Tucker-Drob, 2008).

If measurements across different intervals reflect distinct aspects of change, they might be expected to differ in their patterns of relations with various individual difference characteristics. For example, the age of the participant might be expected to be positively correlated with short-term gains if older adults have less familiarity with testing than younger adults, whereas negative relations of age with longer-term change might be expected if there are age-related declines in the relevant cognitive ability. Both expectations have been supported in previous research as Salthouse and Tucker-Drob (2008) found gains over an interval of approximately one week

were larger among older adults than younger adults, and Rabbitt, Lunn, Wong, and Cobain (2008) found more negative change (smaller gains) across a four-year interval for older adults compared to middle-aged adults.

Because there are theoretical reasons to make opposite predictions regarding its relations with both short-term change and long-term change, another interesting individual difference variable in terms of its relations with cognitive change is general cognitive ability. To illustrate, high-ability individuals might be postulated to exhibit the greatest short-term gains in performance if those gains are reflections of ability-dependent learning, whereas lower-ability individuals would be hypothesized to have the greatest benefits if the additional experience is associated with a reduction in anxiety that was limiting their performance, or with the development of strategies that were not already available to these individuals. Prior research on relations of ability on short-term change has been inconsistent, with some reports of greater short-term gain among individuals with higher levels of general cognitive ability (e.g., Kulik, Kulik, & Bangert, 1984; Rapport, Brines, Axelrod, & Theisen, 1997), some reports of no ability-change relations (e.g., Duff, Callister, Dennett, & Tometich, 2012), and some reports of the greater gains among lower-ability individuals (i.e., Duff et al., 2008; Te Nijenhuis, van Vianen, & van der Flier, 2007). Furthermore, the cognitive reserve hypothesis (Stern, 2003) predicts smaller longitudinal declines among individuals of higher initial ability, but no relations between initial ability and longitudinal change were found in a recent study after controlling influences associated with regression-to-the-mean (Salthouse, 2012a).

Relations between short-term and long-term change are of interest for at least three reasons. One reason is that practice effects over a short interval may have diagnostic significance for the individual's later status. That is, a number of reports have suggested that individuals with the smallest performance gains when a test is repeated after a short interval have a poor prognosis for subsequent cognitive functioning (see Duff, 2012, for a review).

The relation between short-term and longer-term change is also relevant to studies examining effects of a manipulation or intervention across an interval of days to months because it is tempting to assume that the short-term effects are informative about the age-related change that occurs over a period of years or decades. In fact, a study by Zimprich, Hofer, and Aartsen (2004) found a moderate positive correlation between short-term change across three successive trials in a letter coding task and the longer-term change in average

letter coding performance across an interval of approximately three years. However, with this single exception, very little is currently known about the relation between changes occurring across different intervals, and extrapolations from short intervals to longer intervals may only be appropriate if the relations are positive, and at least moderately strong.

Relations between short-term and longer-term change may also be useful in evaluating the viability of the suggestion that gains over short intervals might serve as estimates of the retest effects in longitudinal comparisons (Hoffman, Hofer, & Sliwinski, 2011). As in the prior example, this proposal may only be plausible if the relations between change over short and long intervals are positive, and at least moderately strong.

Although examination of relations between short-term and longer-term change might seem straightforward, dependencies among the change measures need to be considered when examining relations among different types of change. For example, in the terminology of Fig. 1, the 11-to-13 change could be compared with the 11-to-21 change, as recently done for different reasons in Salthouse (2012b, in press-a). However, because the former interval is included in the latter interval, the resulting correlations may be spuriously positive. A less biased comparison involving the changes portrayed in Fig. 1, because each of the assessments in the two sets of changes is separate and at least conceptually distinct, involves the 11-to-12 contrast as an estimate of short-term change and the 13-to-21 contrast as an estimate of longer-term change. Comparisons of these two types of change would indicate whether people who experience the most positive change over an interval of days or weeks (i.e., 11-to-12) are also likely to experience the greatest positive change over an interval of years (i.e., 13-to-21). If this is not the case, it may not be valid to extrapolate findings apparent over short intervals to longer intervals.

To summarize, the three goals of the current project were to investigate influences on cognitive change by comparing change across both short and long intervals, examining individual difference relations with each type of change, and determining the correlations between short-term change and longer-term change. The analyses were based on data from participants in the Virginia Cognitive Aging Project (VCAP; see Salthouse, Pink, & Tucker-Drob, 2008) who had completed at least two ($N = 2017$) or three ($N = 948$) occasions. The data set is uniquely suited to investigate the preceding questions because the participants performed three versions of each test at each occasion in a measurement burst design, cognition was broadly assessed with at least three tests representing each of five cognitive abilities, performance was reliably assessed with composite scores, and the ages of the participants spanned nearly all of adulthood.

2. Methods

2.1. Participants

Characteristics of the participants are reported in Table 1. Although individuals as old as 99 have participated in VCAP, the current analyses are restricted to participants between 18 and 80 years of age to minimize the impact of health-related cognitive impairments that might be more prevalent at older ages. The samples are described in terms of three groups, but

it is important to note that the primary analyses were conducted with age as a continuous, rather than categorical, variable. Some participants only performed a single version of each test on the first session of the first occasion, and performed different tests on the second and third sessions. However, everyone performed all three test versions on the second and third occasions.

The self-identified ethnicity of the participants was primarily white (81%), with about 10% black, and the remainder split among different groups including mixed ethnicity. The testing was conducted between 2001 and 2012, but there were minimal effects of test year on performance (Salthouse, 2013), and thus the data were aggregated across all years. In order to minimize the possibility that the participants had dementia, data from participants with Mini-Mental Status Examination (Folstein, Folstein, & McHugh, 1975) scores less than 24 on the second occasion were excluded.

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 estimated IQ values (see below) indicate that the sample participants performed between .5 and 1 standard deviations above the average of the nationally representative normative samples. The participants in the sample can therefore be inferred to be relatively high functioning, both in terms of years of education and level of cognitive ability.

2.2. Cognitive tests

Sixteen cognitive tests, representing five cognitive abilities, were administered in the same order to all participants. Vocabulary was represented by a 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 set test, and a series completion test. Spatial visualization (space) ability was represented by a spatial relation test, a paper

Table 1
Sample characteristics.

	18–39	40–59	60–80	Age correlation
N @ T1	401/167 ^a	905/393 ^a	711/342 ^a	NA
N @ T2	401	905	711	NA
N @ T3	153	452	343	NA
Age	27.8 (6.9)	50.8 (5.4)	68.8 (5.8)	NA
Prop. female	.66	.72	.62	–.03
Years education	14.7 (2.3)	15.7 (2.6)	16.3 (2.8)	.22*
Self-rated health	2.1 (.8)	2.1 (.9)	2.2 (.9)	.08*
Est. IQ	107.4 (13.6)	110.8 (14.7)	112.1 (13.2)	.11*
PC1 at 11	.35 (1.1)	.08 (1.0)	–.28 (.87)	–.26*
Intervals				
11–12 (days)	5.8 (8.1)	5.6 (5.8)	5.5 (5.9)	–.02
12–13 (days)	5.2 (7.5)	5.0 (6.2)	4.8 (6.1)	–.03
13–21 (years)	2.8 (1.6)	3.0 (1.6)	2.8 (1.3)	–.05
21–22 (days)	5.1 (7.4)	5.9 (8.3)	5.0 (5.4)	–.02
22–23 (days)	5.6 (8.2)	5.8 (7.1)	4.8 (5.9)	–.04
23–31 (years)	3.1 (1.4)	3.1 (1.3)	2.9 (1.2)	–.05
31–32 (days)	5.8 (7.3)	5.4 (6.3)	5.8 (8.4)	.01
32–33 (days)	5.9 (7.4)	6.6 (8.2)	5.7 (7.2)	.01

* $p < .05$

Note. Numbers in parentheses are standard deviations.

^a Refers to the number of participants with alternate versions of the tests on sessions 2 and 3 in the first occasion.

folding test, and a form board test. Episodic memory (memory) was represented by word recall, paired associates, and story (logical) memory tests. Perceptual speed (speed) ability was represented by a digit symbol substitution test, and pattern comparison and letter comparison tests. Details of the tests, including reliabilities and results of factor analyses supporting the hypothesized ability structure, are reported in other publications (Salthouse, 2007; Salthouse & Tucker-Drob, 2008; Salthouse et al., 2008). In order to focus on cognitive abilities rather than specific tests, all of the analyses were conducted on composite scores formed by averaging the z-scores for the three or four (for vocabulary) tests representing each ability.

The three versions of each test were performed in the same order on successive sessions to avoid confounding presentation order with pre-existing characteristics of the individuals. Furthermore, the same order of presentation was used at each occasion (e.g., the version presented at 12 was the same version presented at 22). The composite scores based on different versions were highly correlated across sessions, with correlations ranging from .78 to .91, and thus they can be inferred to represent the same dimensions of individual differences.

A measure of general cognitive ability was derived from the first principal component (PC1) of the 16 cognitive scores at the first session of the first occasion (i.e., 11). It was associated with 41.4% of the variance among the 16 test scores, and it had correlations of .84 with the estimated IQ on the first occasion, and $-.26$ with age.

2.3. Assessment of sample representativeness

In a recent study (Salthouse, *in press-b*) 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 in the sample of 90 adults who performed both batteries. 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.

2.4. Analysis plan

Two steps were carried out to express the scores on the different tests and test versions in comparable units. First, scores on the versions administered in sessions 2 and 3 were adjusted for version differences, and second, all scores were converted to z-scores based on means and standard deviations from the 11 assessment. The version adjustment procedure was based on regression equations from a sample of 90 adults who performed the three versions of the tests in counterbalanced order (Salthouse, 2007), rather than the fixed order that exists in VCAP (which is desirable to treat all individuals the same, and not confound how they are treated with pre-existing characteristics). Specifically, the intercepts and slopes of the regression

equations were used to adjust the scores of every participant on the second and third sessions to remove any order-independent version differences in the means. To illustrate, the intercept and the slope for the equation predicting matrix reasoning scores on the first session from the scores on the second session were 1.86 and .84, respectively. Applying these parameters to an individual with a session 2 score of 12 would result in an adjusted session 2 score of 11.94 (i.e., $1.86 + .84 * 12$).

All of the original and adjusted scores were converted to comparable units by expressing them in z-score units relative to the initial (11) assessments (i.e., each score was subtracted from the corresponding mean at 11 and divided by the standard deviation at 11). Composite scores were then created for each cognitive ability by averaging z-scores for three (or four for vocabulary) tests representing each ability.

Estimates of change in each ability domain were derived by subtracting the first composite score from the second composite score. Because of the direction of the subtraction, positive change values correspond to gains in performance, and negative scores correspond to losses.

There was a considerable amount of missing data because some participants performed different tests on the second and third sessions of the first occasion, and only 47% of the participants had completed a third occasion by the time the data were analyzed. Most of the analyses were therefore carried out with the full information maximum likelihood (FIML) procedure within the AMOS (Arbuckle, 2007) statistical package to deal with the missing data, under the assumption that the data were missing at random. However, it is important to note that the patterns based on analyses of participants with complete data were very similar to those obtained when the FIML algorithm was used to deal with missing data.

Estimates of the means at each assessment, and of the changes from one assessment to the next, were obtained from structural equation models specifying correlations among all of the composite score means, or the composite score differences, for a given ability. By allowing all of the measures to be related to one another, information from measures with complete data was able to influence estimates of measures with missing data. Influences of age and general cognitive ability were estimated with multiple regression models carried out in the AMOS program in which each composite score difference was predicted from age, the measure of general cognitive ability, and their interaction.

3. Results¹

3.1. Interval effects

The rows in the bottom of Table 1 contain the average intervals between successive assessments. The within-occasion intervals (e.g., 11-to-12, 12-to-13, etc.) averaged less than one week, and all of the correlations of these intervals with the magnitude of change in each ability were small (i.e., range from $-.08$ to $.14$), and thus the within-occasion intervals were ignored in subsequent analyses.

¹ Because of the large number of statistical comparisons and the moderately large sample size, a significance level of .01 was used in all statistical tests.

The intervals between the first and the second occasion ranged from less than 1 to more than 11 years, with a mean of 2.9 and a standard deviation of 1.5, and the intervals between the second and third occasions ranged from less than 1 to over 8 years, with a mean of 3.1 and a standard deviation of 1.3. Effects of the intervals between occasions on between-occasion change were investigated in multiple regression analyses with age, interval, and their interaction as predictors of the 13-to-21 and 23-to-31 change in each composite score.² There were significant effects of the interval between the first and second occasions on the 13-to-21 change in the memory and speed composite scores, and of the interval between the second and third occasions on the 23-to-31 change in the memory composite score, in each case in the direction of more negative change at longer intervals. The only significant age-by-interval interaction was with the change from 13-to-21 in the spatial visualization composite score, and it was in the direction of smaller interval effects at older ages. Because significant interval effects were only evident with the memory and speed composite scores, and importantly, only one interaction with age was significant, the effects of between-occasion interval were not considered in subsequent analyses.

3.2. Means across sessions and occasions

For each cognitive ability a structural equation model was specified with correlations among the nine variables corresponding to the composite scores in the three sessions in each of the three occasions. The FIML-estimated means and standard errors at each assessment for participants in three age groups are portrayed in Fig. 2 for the memory composite scores, and Fig. 3 contains comparable information for the other four cognitive abilities. It can be seen that there were large within-occasion (i.e., 11 to 13, 21 to 23, and 31 to 33) increases in the scores on the memory, speed, and space abilities, and sizable between-occasion (i.e., 13-to-21 and 23-to-31) decreases in each ability in at least one of the age groups. Furthermore, the within-occasion changes were generally more positive at older ages, whereas the between-occasion changes were more negative at older ages. The vocabulary data were an exception to this pattern as there was a decrease in performance from the first to the second session (i.e., 11 to 12) for the older adults, but an increase for the younger adults. These patterns are puzzling because they were evident with all four vocabulary tests, and thus are not specific to a particular set of items or to particular tests.

The means and variances for each change measure are reported in the second and third columns of Table 2. Entries in the first row in each set are values for the change from the first assessment on one occasion (e.g., 11) to the first assessment on a subsequent occasion (e.g., 21), roughly corresponding to the contrast in a traditional longitudinal study. The remaining rows contain values for the different components of change. For example, the second row in each set (11–12) contains the values for change from the first to the second session on the first occasion.

It can be seen that all of the estimated variances were significantly greater than zero, indicating that there were significant individual differences in the magnitude of the change. Many of the estimated mean differences were also significantly different from zero, with positive values for most of the short-term (within-occasion) changes and negative values for most of the longer-term (between-occasion) changes. Of particular interest is the contrast between the change from the first session of each occasion (i.e., 11 to 21, and 21 to 31) and the change from the third session of one occasion to the first session of the next occasion (i.e., 13 to 21, and 23 to 31). The latter changes, in which the participants have experience with versions of the tests at each occasion before the assessment of change, were more negative than the former with the memory, speed, and space composite scores, but not with the vocabulary or reasoning composite scores.

3.3. Correlates of changes

Simultaneous regression analyses conducted in AMOS were used to predict each composite score change from age, general cognitive ability (PC1), and the interaction of age and ability, after centering age to minimize collinearity of the interaction term. Quadratic age trends were also examined, but only a few were significant and all were small, and therefore they are not reported.

Longitudinal measurement invariance was examined in a prior study (e.g., Salthouse, 2012c), and the results indicated that the ability factors could be assumed to represent very similar if not identical constructs at both measurement occasions. Earlier studies (e.g., Salthouse, 2011, 2012a) have also found very similar patterns of change relations in analyses involving composite score differences and analyses with latent change models, and that was also the case in the current study. However, because results were not available from all latent change models because of convergence failures, only the results with composite score differences are reported.

The entries in the three columns on the right of Table 2 are estimates of the relations of age and general cognitive ability on each type of composite score difference. Positive coefficients indicate that increased age, or higher levels of general ability, were associated with more positive change, and negative coefficients indicate that greater age, or higher levels of ability, were associated with less positive (more negative) change. Because only a few age-by-ability interactions in Table 2 were significant, and the directions were not consistent, they may have been attributable to chance and are not interpreted.

Although the individual contrasts were not always significant, there was a similar pattern of age relations with memory, speed, reasoning, and space abilities. In each case there were negative relations of age on long-term change, whether the change was measured between the first sessions in each occasion (11-to-21, 21-to-31), or from the third session of one occasion to the first session of the next occasion (13-to-21, 23-to-31). With the exception of space ability, the age relations were small to non-existent on change from the second to the third session on each occasion (i.e., 12-to-13 and 22-to-23). Furthermore, increased age was positively correlated with change from the first to the second session in the second and third occasions (21-to-22 and

² As in other analyses of portions of these data (Salthouse, 2011), the interval effects were significant in each ability domain with the 11-to-21 measure of change.

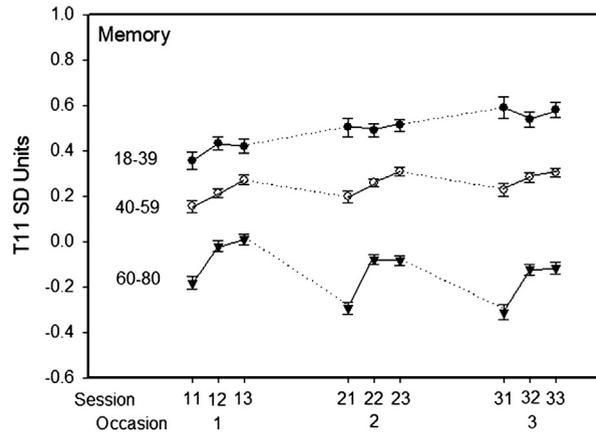


Fig. 2. Estimated means and standard errors for memory composite scores at each assessment among adults in three age groups.

31-to-32), but this was not the case in the first occasion (11-to-12).

With the vocabulary composite score there were also negative relations of age on between-occasion change between

the first sessions of each occasion (11-to-21, 21-to-31), but the relations on change from the third session of one occasion to the first session of the next occasion (13-to-21, 23-to-31) were positive. In contrast to the other abilities, age was negatively

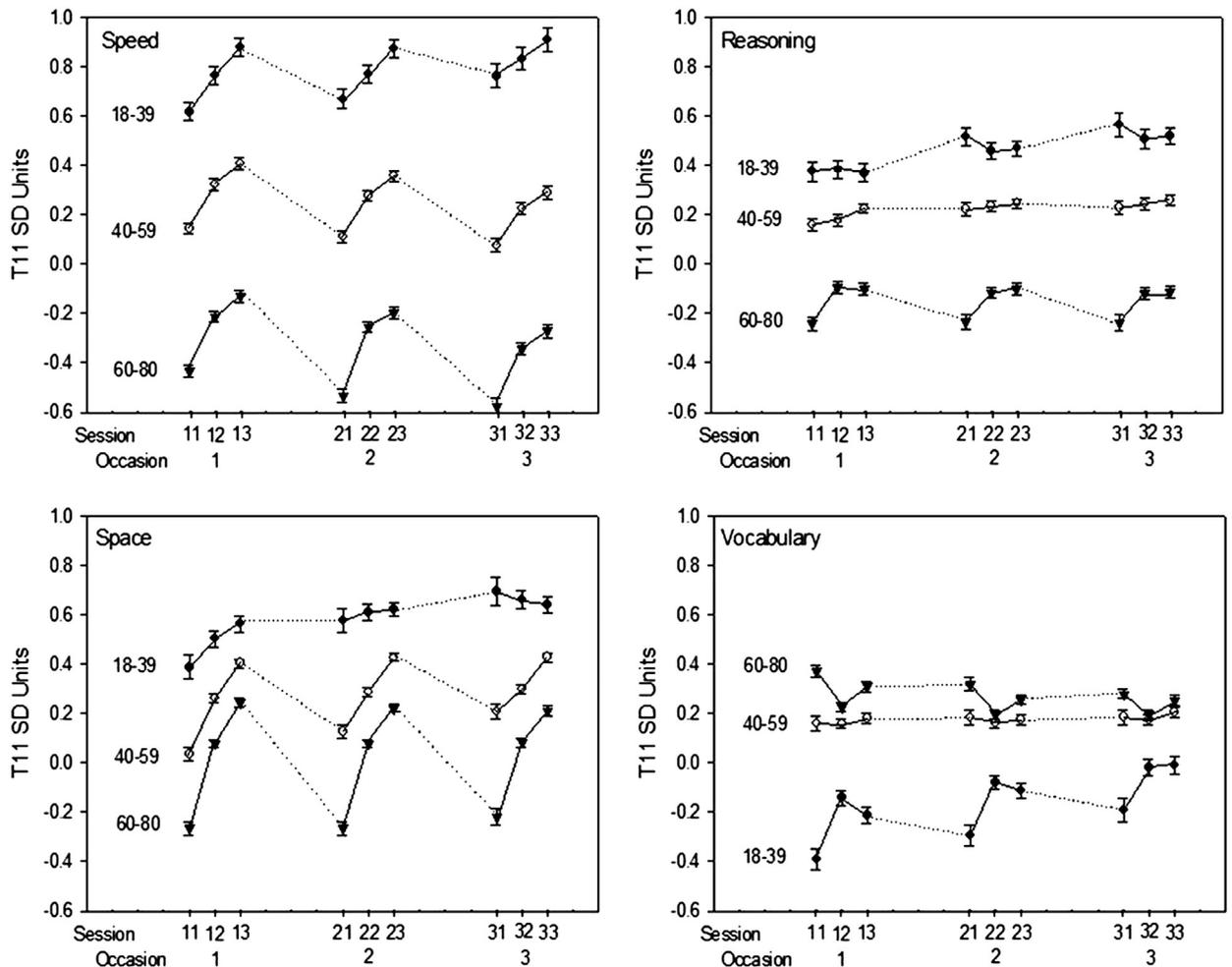


Fig. 3. Estimated means and standard errors for speed, reasoning, space and vocabulary composite scores at each assessment among adults in three age groups.

Table 2

Estimated means and variances of composite score differences, and unstandardized (and standardized) regression coefficients for the relations of age and PC11 on the composite score differences in each ability.

Difference	Mean/variance	Regression predictors		
		Age	PC11	Age * PC11
Memory				
11–21	.01/.18*	-.007 (-.21)*	-.005 (-.02)	-.001 (-.02)
11–12	.15*/.18*	-.001 (-.03)	-.188 (-.41)*	.000 (.02)
12–13	.04*/.10*	.000 (.02)	.001 (.00)	.000 (-.01)
13–21	-.12*/.20*	-.007 (-.24)*	.169 (.34)*	.000 (-.01)
21–31	.01/.24*	-.003 (-.09)	.022 (.05)	.000 (-.02)
21–22	.10*/.16*	.004 (.13)*	-.150 (-.35)*	.000 (.00)
22–23	.03*/.09*	-.001 (-.04)	.017 (.06)	.001 (.06)
23–31	-.10*/.20*	-.005 (-.16)*	.143 (.30)*	-.002 (-.07)
31–32	.07*/.13*	.004 (.15)*	-.148 (-.37)*	.001 (.04)
32–33	.03*/.08*	-.001 (-.06)	.008 (.03)	.001 (.04)
Speed				
11–21	-.05*/.20*	-.004 (-.15)*	-.019 (-.04)	-.002 (-.06)
11–12	.23*/.12*	.001 (.03)	-.046 (-.13)*	-.001 (-.06)
12–13	.11*/.09*	-.001 (-.07)	.003 (.01)	.002 (.13)*
13–21	-.29*/.17*	-.003 (-.13)*	.017 (.04)	-.002 (-.07)
21–31	-.09*/.25*	-.006 (-.20)*	.012 (.02)	-.002 (-.07)
21–22	.19*/.15*	.004 (.15)*	-.059 (-.15)*	.000 (-.02)
22–23	.08*/.09*	-.001 (-.06)	.023 (.08)	.000 (.02)
23–31	-.27*/.21*	-.007 (-.22)*	.073 (.15)*	-.001 (-.02)
31–32	.15*/.12*	.004 (.17)*	-.069 (-.19)*	.000 (.01)
32–33	.07*/.08*	.000 (.01)	.028 (.10)	.001 (.06)
Reasoning				
11–21	.06*/.16*	-.004 (-.14)*	-.051 (-.13)*	.000 (-.01)
11–12	.05*/.13*	.001 (.02)	-.180 (-.45)*	-.003 (-.10)*
12–13	.01/.11*	.000 (-.02)	-.054 (-.16)*	.000 (-.02)
13–21	-.01/.16*	-.004 (-.14)*	.205 (.45)*	.002 (.07)
21–31	.00/.19*	-.002 (-.07)*	.040 (.09)	-.001 (-.03)
21–22	.02/.13*	.003 (.10)*	-.140 (-.36)*	-.001 (-.05)
22–23	.01/.10*	-.001 (-.03)	-.044 (-.14)*	-.001 (-.04)
23–31	-.03/.17*	-.002 (-.07)	.219 (.47)*	.001 (.03)
31–32	.03/.12*	.002 (.09)*	-.168 (-.43)*	.000 (.00)
32–33	.00/.10*	.000 (-.02)	-.027 (-.09)	-.002 (-.09)
Space				
11–21	.08*/.17*	-.004 (-.17)*	.004 (-.02)	-.001 (-.03)
11–12	.27*/.13*	.001 (.04)	-.318 (-.66)*	.000 (-.00)
12–13	.14*/.10*	.002 (.09)	.005 (.02)	.001 (.06)
13–21	-.27*/.19*	-.008 (-.22)*	.326 (.57)*	-.002 (-.04)
21–31	.07*/.19*	-.003 (-.09)*	.018 (-.03)	-.001 (-.03)
21–22	.19*/.16*	.003 (.11)*	-.294 (-.59)*	.000 (-.00)
22–23	.13*/.10*	.002 (.11)*	-.021 (-.07)	.002 (.09)*
23–31	-.25*/.21*	-.007 (-.19)*	.331 (.57)*	-.002 (-.06)
31–32	.13*/.15*	.004 (.13)*	-.316 (-.62)*	.000 (.01)
32–33	.11*/.10*	.003 (.17)*	.018 (.06)	.001 (.06)
Vocabulary				
11–21	.01/.09*	-.004 (-.19)*	-.016 (-.05)	.000 (.02)
11–12	.03/.12*	-.014 (-.47)*	-.282 (-.63)*	.000 (-.02)
12–13	.02/.09*	.003 (.18)*	.035 (.12)*	.000 (.02)
13–21	-.04*/.13*	.007 (.25)*	.235 (.55)*	.000 (.02)
21–31	.00/.13*	-.005 (-.21)*	.012 (.03)	.000 (.01)
21–22	-.01/.12*	-.012 (-.44)*	-.244 (-.57)*	.000 (.02)
22–23	.02*/.08*	.003 (.17)*	.031 (.11)*	.000 (-.01)
23–31	.01/.15*	.006 (.21)*	.214 (.50)*	-.001 (-.04)
31–32	-.00/.12*	-.010 (-.38)*	-.241 (-.57)*	.000 (-.01)
32–33	.04*/.06*	.002 (.14)*	.020 (.08)	.000 (.03)

Note. Numbers in parentheses are standardized coefficients.

* $p < .01$.

related to change from the first to the second session of each occasion, indicating greater within-occasion gains in vocabulary at younger ages.

There were negative relations of general cognitive ability (PC1) with change from the first to the second session on each occasion (i.e., 11-to-12, 21-to-22, and 31-to-32) in each ability

domain, indicating smaller short-term gains among individuals at higher ability levels. The change from the second to the third session in each occasion was small, and only weakly related to overall ability. Most of the relations of cognitive ability to between-occasion change based on the contrast of first sessions in each occasion (i.e., 11-to-21 and 21-to-31) were

not significant, but there were positive relations of cognitive ability with between-occasion change based on the contrast of the third session in one occasion to first session in the next occasion (i.e., from 13-to-21, and from 23-to-31).

Fig. 4 portrays the estimated changes in the memory composite scores across three age groups in the top panel, and across three ability groups in the bottom panel. The left two panels contain the changes from the 11 to the 21 assessments, and the right panels contain the assessments from the 21 to 31 assessments. The first value in each set (black bar) is the change between the initial assessment in the two occasions (i.e., 11-to-21 and 21-to-31), the second value in each set is the change from the first to the second session (i.e., 11-to-12 and 21-to-22), and the third value is change from the second to the third session (i.e., 12-to-13 and 22-to-23). The fourth value (white bar) is the change from the third session of one occasion to the first session of the next occasion (i.e., 13-to-21 and 23-to-31).

The top two panels in Fig. 4 portray results for three age groups when general cognitive ability (as indexed by the PC1) was controlled at the average level in the sample. Note that the within-occasion change from the first to the second session (gray bars) was similar at each age on the first occasion, but was greater at older ages on the second occasion. Furthermore, both the between-occasion change based on the first assessments in each occasion (black bars), and that based on the change between the third session of one occasion and the first session on the following occasion (white bars), were more negative at older ages.

The bottom two panels in Fig. 4 portray estimates of memory change when the sample was divided into three ability groups, with age controlled at the average value in the sample. The ability groups were created by dividing the sample on the basis of the PC1 variable into participants of high ability (i.e., PC1 above .5), participants of moderate ability (i.e., PC1 between $-.5$ and $.5$), and participants of low ability (i.e., PC1 below $-.5$). It can be seen that the between-occasion changes for the contrast of first session scores in each occasion (black bars) were all relatively small, but that the change from the third session of one occasion to the first session of the next occasion (white bars) was more negative among individuals at lower ability levels. The trends in the gray bars representing change from the first to the second assessment within each occasion indicate greater short-term gains in individuals of low ability than in those of higher overall ability.

3.4. Relations between short term and longer term change

Relations between short-term and longer-term change were examined with regression analyses in which the 11-to-12 (short-term, within-occasion) change, age, and the interaction of short-term change and age were used as predictors of the 13-to-21 (longer-term, between-occasion) change. Parallel analyses were also conducted with the 21-to-22 contrast as the measure of short-term change and the 23-to-31 contrast as a measure of longer-term change, and results from both sets of analyses are reported in Table 3.

Inspection of the entries in Table 3 reveals that there were significant negative relations between within-occasion change and between-occasion change in each cognitive domain.

Furthermore, there were negative effects of age on between-occasion change for all cognitive abilities except vocabulary. The only significant interaction of age and short-term change on longer-term change occurred with vocabulary change from 23-to-31, which was in the direction of a less negative relation of 21-to-22 change with 23-to-31 change at older ages. Contrary to the expectation that change over short intervals might resemble change over longer intervals, the negative correlations in Table 3 indicate that individuals with the greatest short-term gains from the first to the second assessment in one occasion tended to have the largest longer-term losses from the third assessment of one occasion to the first assessment of the next occasion. Furthermore, with the possible exception of change from the second to the third occasion with vocabulary ability, there is no evidence that this pattern varied as a function of age.

4. Discussion

In a typical longitudinal study change can be hypothesized to consist of an unknown mixture of change attributable to increased experience and shift in the relevant construct because the two types of influences on change cannot be distinguished. However, various components of change, which might be differentially sensitive to these influences, can be distinguished when multiple assessments are available at each occasion, as in the current measurement burst design.

Although the measurement burst design portrayed in the bottom panel of Fig. 1 is valuable for distinguishing hypothesized components of change, it is important to note that comparisons of this design with a traditional longitudinal design (represented in the top panel of Fig. 1) will be confounded if there are reactive effects of the 12 and 13 assessments on the longitudinal change from 11-to-21. In fact, Salthouse (in press-a) recently reported more positive 11-to-21 change in participants who performed different versions of the same tests, as opposed to different tests, on the second (12) and third (13) sessions of the first occasion. Because the magnitude of longitudinal change was affected by additional assessments involving the same tests, it may not be possible to distinguish components of change without introducing reactive effects in a traditional longitudinal study.

One noteworthy finding was that the gains from the first to the second assessment on the first occasion were similar among adults of different ages, but that the short-term gains from the first to the second assessments on both the second and third occasions were greater at older ages. In addition, increased age was associated with more negative longer term (between-occasion) change, both in the contrast of the scores at the first assessment in each occasion (i.e., 11-to-21 and 21-to-31), and in the contrast of the scores from the last assessment in one occasion to the first assessment in the subsequent occasion (i.e., 13-to-21 and 23-to-31).

The pattern just described suggests that two factors may be involved in adult age differences in longitudinal change. One factor is a possible decline in level of ability, as reflected in both types of between-occasion changes. The second factor is loss of the short-term gains associated with test-specific skills and strategies or construct-irrelevant factors, which can be inferred from the greater within-occasion gains (i.e., 21-to-22 and 31-to-32) on the second and third occasions at older ages. That is, the small within-occasion gains on the second and third

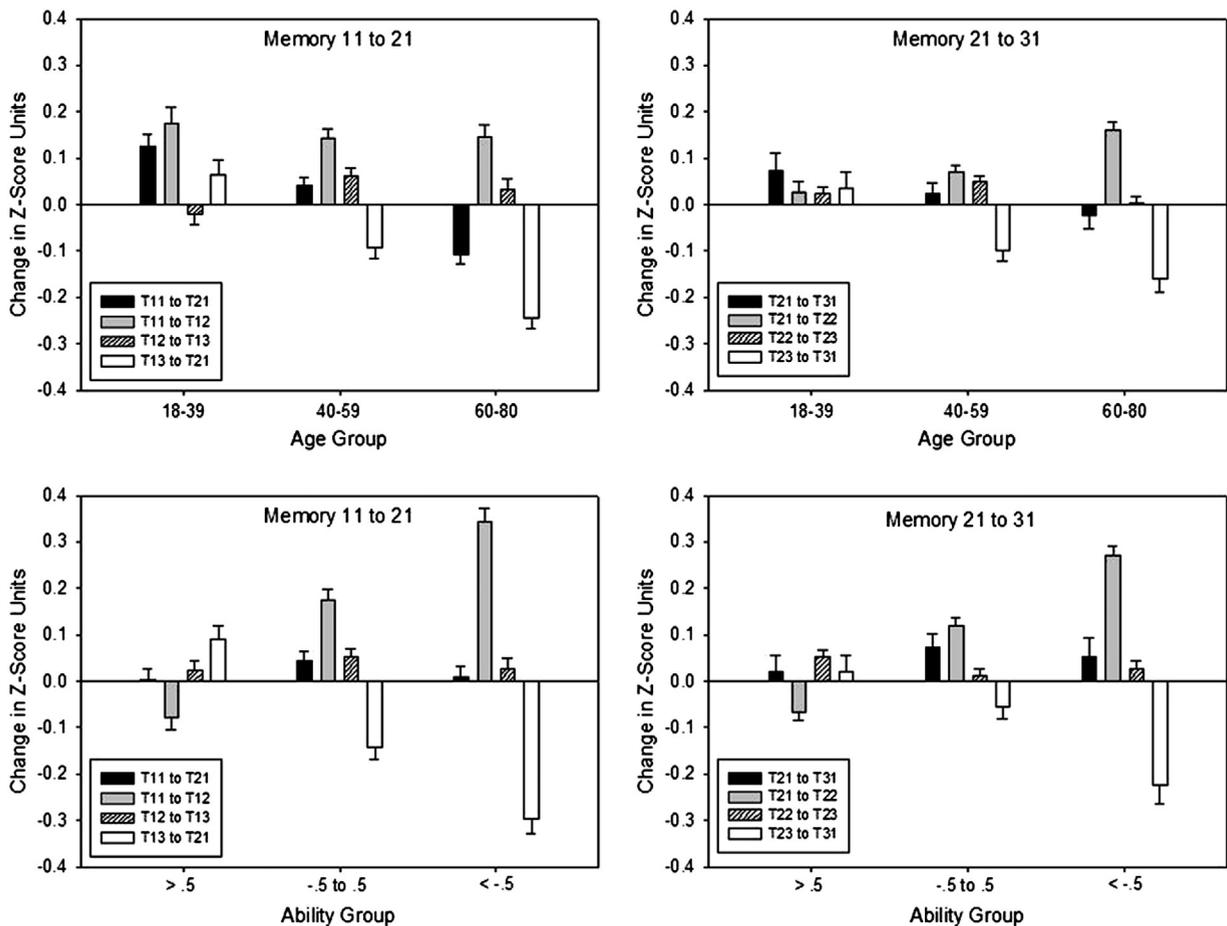


Fig. 4. Between-occasion and within-occasion changes in memory composite scores in three age groups (top panels) and in three ability groups (bottom panels). Panels on the left portray change from the 11 to 21 assessments, and panels on the right portray change from the 21 to 31 assessments. Note that the ability groups are ordered from highest ($PC11 > .5$) on the left to lowest ($PC11 < -.5$) on the right. The values are based on FIML estimates controlling general cognitive ability in the top two panels and controlling age in the bottom two panels.

occasions at younger ages are consistent with these individuals having experienced less loss of whatever contributes to within-occasion gains over the longitudinal (between-occasion) interval than the older participants. Both of these hypothesized factors may be contributing to cognitive change in many longitudinal studies, but they are only distinguishable when change can be assessed both within and across occasions, as in a measurement burst design.

Individuals with lower levels of general cognitive ability had greater short-term gains at each occasion than high-ability individuals, indicating that at least with this sample of relatively healthy adults, the benefits of additional test experience over an interval of days to weeks were greater for individuals with the lowest initial levels of ability. Although there was a positive relation of general cognitive ability with change from the last assessment in one occasion to the first assessment in the subsequent occasion (i.e., 13-to-21 and 23-to-31), this may be a consequence of the greater short-term (i.e., 11-to-12, and 21-to-22) gains at higher ability levels because there was no relation of ability when change was evaluated between the first assessments in each occasion (i.e., 11-to-21 and 21-to-31).

As noted earlier, several studies have reported greater short-term gains among higher-ability individuals, whereas

the opposite pattern was evident in this study. One potential reason for the discrepant results is that participants in some of the earlier studies may have been relatively low functioning, such that they had impairments in learning that were manifested in the lack of short-term practice effects. Another possible explanation of the discrepancy concerns the nature of the repeated tests, which consisted of different items in the three sessions at each occasion in the current study, but identical items at each assessment in several of the previous studies. It is therefore conceivable that low-ability individuals experience the greatest benefits when the successive tests involve different items, and most of the gains are associated with reduction in anxiety and/or acquisition of an effective strategy, whereas higher-ability individuals might have greater benefits when the tests involve identical items and memory of the earlier experience contributes to the practice gains. Results from a study by Salthouse and Tucker-Drob (2008) are consistent with this interpretation as some of the participants in that study performed identical versions of the tests on the first and second sessions of the first occasion, and the relations of general cognitive ability with the 11-to-12 change in those individuals were positive, rather than negative as in the current study.

Table 3

Unstandardized (and standardized) regression coefficients for the prediction of between-occasion change (13-to-21 or 23-to-31) from within-occasion change (11-to-12 or 21-to-22).

	Between occasion change	
	13-to-21 change	23-to-31 change
Memory		
Within-occasion change	-.22 (-.20)*	-.20 (-.17)*
Age	-.01 (-.33)*	-.01 (-.16)*
Within* age	.00 (.03)	-.00 (-.04)
Speed		
Within-occasion change	-.15 (-.12)*	-.24 (-.20)*
Age	-.00 (-.15)*	-.01 (-.17)*
Within* age	.00 (.03)	-.00 (-.05)
Reasoning		
Within-occasion change	-.26 (-.23)*	-.33 (-.27)*
Age	-.01 (-.22)*	-.00 (-.11)*
Within* age	-.00 (-.02)	.00 (.02)
Space		
Within-occasion change	-.63 (-.53)*	-.55 (-.47)*
Age	-.01 (-.27)*	-.01 (-.21)*
Within* age	.00 (.04)	.01 (.07)
Vocabulary		
Within-occasion change	-.51 (-.54)*	-.50 (-.49)*
Age	-.00 (-.07)	-.00 (-.07)
Within* age	.00 (.07)	.01 (.09)*

* $p < .01$.

The negative relations between short-term (within occasion) and longer-term (between-occasion) change reported here raise questions about the viability of proposals to use short-term change as an estimate of retest effects in longitudinal designs (Hoffman et al., 2011), and are inconsistent with the results of Zimprich et al. (2004), who reported a moderate positive correlation. The reasons for this latter discrepancy are not clear, but it is worth noting that the earlier study was based on a single cognitive measure and only one of three available short-term changes was significantly related to the longer-term change. In contrast, the measures in the current study were composite scores or latent variables formed from three or four separate tests, and the negative relations between short-term change and longer-term change were evident in all five cognitive ability domains.

Because the current study involved multiple measures of change, it is useful to consider properties of alternative measures of longitudinal (across-occasion) change. The contrast between the first assessments in each occasion (as in 11–21, somewhat analogous to the traditional longitudinal comparison) is not ideal because it can be assumed to involve a mixture of experience effects associated with prior testing and possible changes in ability. The contrast between the third assessment in one occasion and the first assessment in the next occasion (as in the 13–21 contrast) allows test experience effects to be minimized by the presence of the 11 and 12 assessments, as advocated in the dual baseline procedure. However, a disadvantage of this contrast is that it neglects experience effects after the first occasion, which are evident in the gains from the first to the third session on both the second and third occasions.

Experience effects at each occasion might be considered by evaluating change across the same sessions at each occasion. That is, change on the first session at each occasion (11–21) could be compared with change on the second (12–22) and third (13–23) sessions (cf. Salthouse, 2012b). Because later

sessions involve progressively more short-term experience, change on later sessions might be less affected by non-ability influences than change on the initial session.

Another approach to dealing with multiple measures at each occasion is to ignore the ordering of the assessments within each occasion, and either average the three measures, or use them as manifest indicators of a latent construct (Salthouse & Nesselrode, 2010). Although methods based on aggregation do not evaluate within-occasion change, they minimize its influence in the evaluation of between-occasion change by combining measures across different amounts of experience. Another advantage of aggregation is that the resulting measures will generally be more reliable than any individual measures of change.

In summary, there are at least four important implications of the current results. One implication is that cognitive change is complex, and not easily characterized. That is, change in the current study varied according to the cognitive domain (e.g., memory, vocabulary, etc.), the length of the interval between assessments (i.e., days or years), and both the age and overall cognitive ability level of the individual.

A second implication is that effects of test experience can be distinguished from other contributors to change with a measurement burst design if within-occasion change is assumed to primarily reflect non-ability influences, whereas across-occasion change is more likely to reflect ability influences (in addition to other influences). Furthermore, analyses of within-occasion change on successive occasions may be informative about the contributions of different types of influences on longitudinal change.

A third implication of the current results is that short-term change does not appear to be a good proxy for what might happen over longer intervals, or as an indicator of subsequent change. Because there was a negative correlation between the types of two changes, the factors involved in short-term change, at least over a period of days to weeks, appear to be distinct from the factors involved in longer-term change occurring over a period of years. More research is needed with intervals ranging from weeks to years to determine the precise relation among changes across different time periods, but the results of this study suggest that patterns apparent across intervals of days or weeks may have weak or even negative relations to patterns occurring across intervals of years.

Finally, the results of this study imply that initial assessments of cognitive functioning could be considered relatively unfair to older individuals, and to lower ability individuals. That is, because people in these categories experience greater gains from the first to a second assessment, the best estimate of the true level of ability of these individuals may only be available after they have had some initial exposure to the tests and the testing situation.

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.

References

- Arbuckle, J. L. (2007). *AMOS (Version 7) [Computer Program]*. Chicago: SPSS.
- Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., et al. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*, *20*, 517–529.
- Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, *27*, 248–261.
- Duff, K., Beglinger, L. J., van der Heiden, S., Moser, D. J., Arndt, S., Schultz, S. K., et al. (2008). Short-term practice effects in amnesic mild cognitive impairment: Implications for diagnosis and treatment. *International Psychogeriatrics*, *20*, 986–999.
- Duff, K., Callister, C., Dennett, K., & Tometich, D. (2012). Practice effects: A unique cognitive variable. *The Clinical Neuropsychologist*, *26*, 1117–1127.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.
- Hoffman, L., Hofer, S. M., & Sliwinski, M. J. (2011). On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: A simulation study. *Psychology and Aging*, *26*, 778–791.
- Kulik, J. A., Kulik, C. -L. C., & Bangert, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, *21*, 435–447.
- McCaffrey, R. J., & Westervelt, H. J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, *5*, 203–221.
- Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palermo (Eds.), *Views of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.
- Rabbitt, P., Lunn, M., Wong, D., & Cobain, M. (2008). Age and ability affect practice gains in longitudinal studies of cognitive change. *Journal of Gerontology: Psychological Sciences*, *63B*, P235–P240.
- Rappport, L., Brines, D. B., Axelrod, B. N., & Theisen, M. E. (1997). Full scale IQ as mediator of practice effects: The rich get richer. *The Clinical Neuropsychologist*, *11*, 375–380.
- 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. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*, 507–514.
- Salthouse, T. A. (2011). Effects of age on time-dependent cognitive change. *Psychological Science*, *22*, 682–688.
- Salthouse, T. A. (2012a). Does the direction and magnitude of cognitive change depend on initial level of ability? *Intelligence*, *40*, 352–361.
- Salthouse, T. A. (2012b). Robust cognitive change. *Journal of the International Neuropsychological Society*, *18*, 749–756.
- Salthouse, T. A. (2012c). Does the level at which cognitive change occurs change with age? *Psychological Science*, *23*, 18–23.
- Salthouse, T. A. (2013). Within-cohort age differences in cognitive functioning. *Psychological Science*, *24*, 123–130.
- Salthouse, T. A. (in press-a). Effects of first occasion test experience on longitudinal cognitive change. *Developmental Psychology*.
- Salthouse, T. A. (in press-b). Evaluating the correspondence of different cognitive batteries. *Assessment*.
- Salthouse, T. A., & Nesselroade, J. R. (2010). Dealing with short-term fluctuation in longitudinal research. *Journal of Gerontology: Psychological Sciences*, *65B*, 698–705.
- 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.
- Stern, Y. (2003). The concept of cognitive reserve: A catalyst for research. *Journal of Clinical and Experimental Neuropsychology*, *25*, 589–593.
- Te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, *35*, 283–300.
- Van Gorp, W., Lamb, D., & Schmitt, F. (1993). Methodologic issues in neuropsychology research with HIV-spectrum disease. *Archives of Clinical Neuropsychology*, *8*, 17–33.
- Wechsler, D. (2008). *WAIS-IV: Administration and scoring manual*. San Antonio, TX: Pearson.
- Zimprich, D., Hofer, S. M., & Aartsen, M. J. (2004). Short-term versus long-term longitudinal changes in processing speed. *Gerontology*, *50*, 17–21.