Published in final edited form as:
Assessment. 2012 December ; 19(4): 494–501. doi:10.1177/1073191112438744.

Psychometric Properties of Within-Person Across-Session Variability in Accuracy of Cognitive Performance

Timothy A. Salthouse[1]
[1]University of Virginia, Charlottesville, VA, USA

Abstract

Although most psychological assessments are based on measures related to an individual's average level of performance, it has been proposed that measures of variability around one's average may provide unique individual difference information and have clinical significance. The current study investigated properties of within-person variability in measures of performance accuracy in a sample of more than 1,700 healthy adults. Contrary to what has been reported with measures of within-person variability in reaction time, measures of within-person variability in performance accuracy from different cognitive tests had weak correlations with one another, very low stability across time, and near-zero correlations with longitudinal change in cognitive abilities.

Keywords

aging; cognition; variability

In recent years, there has been a great deal of interest in within-person, or intraindividual, variability as a potentially important dimension of individual differences in cognitive functioning (see Hultsch, Strauss, Hunter, & MacDonald, 2008, for a review). Measures of within-person variability have been found to differ as a function of normal aging (e.g., Der & Deary, 2006; Hultsch, MacDonald, & Dixon, 2002; Nesselroade & Salthouse, 2004; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005), clinical condition (e.g., Adams, Roberts, Milich, & Fillmore, 2011; Bleiberg. Garmoe, Halpern, Reeves, & Nadler, 1997; Burton, Strauss, Hultsch, Moll, & Hunter, 2006; Christensen et al., 2005; Dixon et al., 2007; Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000; Strauss, Bielak, Bunce, Hunter, & Hultsch, 2007; Stuss, Murphy, Binns, & Alexander, 2003), subsequent cognitive decline (e.g., Bielak, Hultsch, Strauss, MacDonald, & Hunter, 2010; Lovden, Li, Shing, & Lindenberger, 2007; MacDonald, Hultsch, & Dixon, 2003), and time until death (e.g., Deary & Der, 2005; MacDonald, Hultsch, & Dixon, 2008; Shipley, Der, Taylor, & Deary, 2006). In addition, measures of within-person variability have been reported to be stable over short intervals (e.g., Hultsch et al., 2000; Nesselroade & Salthouse, 2004; Rabbitt, Osman, Moore, & Stollery, 2001; Saville et al., 2011) and to have a coherent structure in factor analyses (e.g., Hultsch et al., 2000, 2002; Li, Aggen, Nesselroade, & Baltes, 2001; Strauss et al., 2007). Taken together, these properties suggest that measures of within-person variability

Corresponding Author: Timothy A. Salthouse, Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA salthouse@virginia.edu.

Author's Note: 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.

Declaration of Conflicting Interests: The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

can provide valuable information about an individual's cognitive status. That is, measures of variability in cognitive functioning could be an early indicator of impending change and might serve as a unique marker of certain clinical conditions.

Several characteristics are common to much of the prior research investigating within-person variability. First, although there are exceptions (e.g., Hultsch et al., 2000; Li et al., 2001), most of the studies have investigated variability in reaction to time or other speeded tasks. Second, a large proportion of the studies have focused on short-term variability across trials within a single session rather than longer term variability across different sessions (but see Hultsch et al., 2000; Nesselroade & Salthouse, 2004; Rabbitt et al., 2001; Salthouse & Berish, 2005; Stuss et al., 2003). And third, the majority of the studies involved adults older than about 60 years of age (but see Der & Deary, 2006; Williams et al., 2005). These characteristics lead to questions as to whether within-person variability is primarily meaningful as an individual differences measure when it is based on speeded tasks, when it reflects momentary rather than day-to-day fluctuations, and when the research participants are older adults.

A different, and in some respects more surprising, form of variability is the variation in the accuracy of performance of the same cognitive task across sessions separated by days or weeks. Because cognitive abilities are generally considered to be relatively stable traits, this form of variability is often assumed to be very small. However, recent research has revealed that short-term variability in level of cognitive performance can be substantial (e.g., Salthouse, 2007; Salthouse, Nesselroade, & Berish, 2006). The goal of the current project was to investigate properties of this type of within-person variability to determine if it has some of the same characteristics as within-person variability in reaction time (RT) tasks and whether it may represent a meaningful dimension of individual differences. In particular, the magnitude, stability, and correlations of within-person variability in accuracy of cognitive performance were examined in 16 cognitive tests.

The analyses were conducted on data from the Virginia Cognitive Aging Project (VCAP), which is a mixed cross-sectional and longitudinal study involving a measurement burst design in which different versions of 16 cognitive tests were performed in each of three sessions. Participants in VCAP span a wide age range, and therefore separate analyses were reported for adults aged 60 to 95 years and 18 to 59 years. The former group is typical of most prior research, and the latter group allows the phenomenon to be examined at younger ages.

To summarize, the current study was designed to investigate the following properties of within-person across-session variability in accuracy of cognitive performance: magnitude (i.e., within-person variability relative to between-person variability), structure (i.e., interrelations of measures of within-person variability in different variables), longitudinal stability (i.e., correlations of within-person variability from Time 1 [T1] to Time 2 [T2]), and correlations of within-person variability with change in cognitive abilities from T1 to T2.

Method

Sample

The three-session measurement burst at T1 was completed by 1,725 adults, with 579 of them returning for a second measurement burst after an interval averaging about 2.3 years. Characteristics of the initial sample and of the sub-sample with longitudinal data are presented in Table 1. Note that the longitudinal participants had slightly higher scaled scores in word recall, but in other respects they were similar to the total sample.

Participants were recruited from newspaper advertisements and referrals from other participants. Approximately 81% were Caucasian, about 10% African American, and the remaining distributed across other ethnicities or reporting more than one ethnicity. Most participants performed the sessions at the same time of day, but this was not always the case as the appointments were arranged to accommodate to the participants' schedules.

Cognitive Tests

Cognitive functioning was assessed with 16 tests selected to reflect 5 cognitive abilities (i.e., vocabulary, inductive reasoning, spatial visualization, episodic memory, and perceptual speed). The appendix contains a brief description of the tests and their sources. Most of the test versions had internal consistency and test–retest reliabilities of .7 or greater, and loadings of .7 or greater on their respective ability factors in confirmatory factor analyses of these 16 tests (i.e., Salthouse, 2007; Salthouse, Pink, & Tucker-Drob, 2008; Salthouse & Tucker-Drob, 2008).

The tests were administered in the same order in each session, but different versions of the tests were performed in each of the three sessions. Because the test versions could differ in mean performance, which would preclude direct comparisons across versions, the three versions were administered in a counterbalanced order to a separate sample of 90 adults between 20 and 79 years of age to determine average performance in each version without confounding version with sequence (Salthouse, 2007). Regression equations in this sample were used to predict performance in the original test version from scores on 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 sequence-independent version differences in means. Detailed information on this calibration sample, including means and standard deviations (*SD*s) in each session, is reported in Salthouse (2007).

Analyses reported in Salthouse and Nesselroade (2010) revealed that the average accuracy of performance increased across the three sessions within each measurement occasion. Because any systematic linear trends could inflate the estimates of within-person variability, and confound them with short-term learning, they were removed with regression equations applied to the data of individual participants, and then *SD*s of the residuals were used as the index of within-burst variability for each variable.

Results[1]

As noted above, regression equations were used to adjust the scores on the second and third test versions to have the same expected means as the scores on the first version. All adjusted scores for each variable were then converted into *z*-score units based on the distribution of scores on the first session of the first occasion (T11) for that variable to express scores on different test versions in the same units. Regression analyses were also conducted on each participant's data relating test score to session number, with the *SD*s of the regression residuals used as the measure of within-person variability in each test for each participant. Note that two separate sets of regression equations were applied to the data, one set (which was the same for all participants) to equate the difficulty of the different test versions and a second set (consisting of different equations for each participant) to remove individual-specific across-session practice effects.

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

Magnitude of Standard Deviation

The averages of the within-person *SD*s for each cognitive variable in the two age-groups after the adjustments described above are reported in Table 2. Across the 16 tests, the median *SD* at the first (T1) occasion was .23. Because the T11 scores were in *z*-score units, the between-person *SD*s were very close to 1. The average within-person variability can therefore be inferred to be about 23% of the variation across people on the first assessment.

Another method of evaluating the magnitude of within-person variability is to contrast it with the magnitude of cross-sectional age differences in the sample (Salthouse, 2007; Salthouse et al., 2006). Slopes of regression equations relating T11 scores to age for the variables with negative age relations (i.e., all variables except the vocabulary variables) ranged from −.015 to −.033 *SD* per year, with a median of −.026 *SD* per year. Dividing the median within-person variability (.23) by the median annual cross-sectional difference (−.026) indicated that the within-person variability for these variables was equivalent to about 8.8 years of cross-sectional age difference. The median within-person variability and the number of years of cross-sectional age difference were smaller than in prior studies (i.e., Salthouse, 2007; Salthouse et al., 2006), likely because the linear across-session trends were removed before computing within-person variability.

Inspection of the values in Table 2 reveals that the magnitude of within-person variability in each test was similar in the two age-groups and the effect sizes in *d* units for the group difference were all relatively small. Moreover, this was also true at both T1 and T2 for the longitudinal participants, whose data are presented in the bottom panel of Table 2. Correlations were also examined between age and within-person variability in the entire sample. All the correlations were relatively small, with a range from −.14 to .08 and a median of .01. Although within-person variability in cognitive performance was moderately large compared with both between-person variability and cross-sectional age differences, there was no evidence that these measures of within-person variability are related to the age of the participant.

Structure

An exploratory analysis was conducted to determine if there was structure in the measures of within-person variability. A principal components analysis on the T1 within-person *SD*s revealed that the first component was associated with 9.5% of the variance in the 18- to 59-year age-group and that the cumulative percentage of variance after two components was 17.3%. Corresponding values in the 60- to 95-year age-group were 9.1% and 16.9%, respectively.

For comparison purposes, similar analyses were conducted on the T11 scores representing average performance. In the 18 to 59 age-group, the first component was associated with 51.6% of the variance, and the cumulative percentage with two components was 63.7%. Corresponding values in the 60 to 95 age-group were 39.8% and 51.5%, respectively. The results with the T11 scores replicate the familiar pattern with cognitive test scores, but the weak interrelations with the measures of within-person variability in both age-groups suggest that they have little or no structure and weak relations with one another.

Stability

Correlations across the longitudinal interval were computed for the first session means (i.e., T11 and T21) and for the measures of within-subject variability at T1 and T2. It can be seen in Table 3 that these stability coefficients were moderately high for the Session 1 scores, with medians of .72 and .71, respectively, in the 18 to 59 and 60 to 95 age-groups. However, the correlations were very low for the within-person *SD*s, with medians of only .07 and .08 in the two groups.

Correlations

A final set of analyses examined whether within-person variability at the first occasion (T1) predicted cognitive change from T1 to T2. A latent change model (Ferrer & McArdle, 2010) as portrayed in Figure 1 was used in these analyses. The variables reflecting cognitive ability (e.g., word recall, paired associates, and logical memory for memory ability) are represented by squares, and the latent level (L) and latent change (C) constructs are represented by circles. Of primary interest in the current context are the relations between the measures of within-person variability for a given cognitive test variable (represented in the box at the top of the figure) and the latent change construct (the circle labeled C). Advantages of latent change models for the analysis of change are that change is evaluated in terms of latent constructs that theoretically have no measurement error and that all available data can be used in the analyses with the full-information maximum likelihood algorithm.

A total of 65 (out of 160, consisting of 16 tests related to each of five cognitive abilities in the two age-groups) relations were significant in the prediction of the level parameter in these latent change analyses. Most of the relations were negative, indicating that greater within-person variability was associated with lower levels of cognitive ability. However, only 6 of 160 relations were significant ($p < .01$) in the prediction of the latent change parameter, and they were distributed across different combinations of within-person variability measures and cognitive abilities. These results therefore provide little evidence of a systematic relation of the measures of within-person variability with longitudinal change in cognitive ability.

Discussion

The results reported above indicate that across-session within-person variability in accuracy of cognitive performance is moderately large, as it is almost one fourth the magnitude of the between-person variability in average performance in a given session and is equivalent to nearly 9 years of cross-sectional age difference. However, this type of variability appears to be unsystematic because the correlations with measures of within-person variability in other cognitive tests were very small, with little evidence of structure among the measures; the measures had almost no across-time stability; and the correlations with longitudinal change in cognitive abilities were all very small. Moreover, in each of these respects the pattern was very similar in independent samples of adults between 60 and 95 years of age and between 18 and 59 years of age.

Why is the pattern of within-person variability in accuracy of performance different from that reported with measures of within-person variability in RT, in which the measures of variability have been reported to have moderate stability and significant relations to one another and to other types of variables? Both methodological and substantive factors may be involved. For example, RT variability is sensitive to a few very slow RTs, and in some studies the elimination of extreme scores may have been incomplete because outliers were identified on the basis of group means rather than means of individual participants. It is also possible that in some studies the measure of variability might not have been independent of the mean, and therefore relations with measures of within-person variability may have indirectly reflected relations with the mean. To illustrate, Hultsch et al. (2000) attempted to remove the effects of mean RT by partialing the effects associated with group membership (and occasion and trial effects) before computing *SD*s of the residuals, but these “purified residuals” still had substantial correlations (ranging from .54 to .94) with mean RT. There is currently no consensus on the ideal method of controlling influences of the mean, and complicated methods may be needed because Schmiedek, Lovden, and Lindenberger. (2009) recently reported that the relations between mean and variance can differ across individuals. Whatever method is used, however, before interpreting relations involving within-person variability, it is important to verify independence of mean and variability empirically because within-person variability may not be a unique dimension of individual differences if it is not independent of the mean.

Among the possible substantive reasons for the differences across studies are the time frame over which variability was assessed and the nature of the dependent variable. That is, variability across trials within a single session may reflect the ability to maintain attention over a brief period of time, whereas variability across days may reflect fluctuations in mood or in one's general state. Another possibility is that RT and other measures of speeded processing are simply more sensitive than measures of accuracy of performance, such that subtle aspects of variability are more detectable with RT measures than with accuracy measures. For example, measures of within-person variability with RT can be computed across a few trials, whereas many trials must be aggregated to obtain a sensitive measure of accuracy, and even more are needed to allow within-person variability to be computed.

It should be noted that although not measured in units of time, three of the cognitive tests in this project were designed to assess perceptual speed (i.e., digit symbol, letter comparison, and pattern comparison). However, the results in Tables 2 and 3 indicate that the patterns for these measures were very similar to the others, and therefore the most relevant distinction regarding differential properties of within-person variability may be between RT and other measures of cognitive performance, not between speed and accuracy of performance.

Regardless of the reasons for the differences between RT and other measures of performance, it is important to recognize that other studies have reported results similar to those in this study. For example, Hultsch et al. (2000) noted that their measures of within-person variability of memory accuracy were less sensitive to clinical group membership than RT measures of within-person variability, and Li et al. (2001) reported very low values of stability and reliability of within-person variability for measures of memory accuracy. Furthermore, with data from a subset of the current sample, Salthouse (2007) reported reliabilities of within-person variability based on correlations of within-person *SD*s from odd-numbered and even-numbered items. The median reliabilities across the cognitive variables in the two studies, respectively, were .26 and .26, compared with median reliabilities of .94 and .92 for the mean scores in the same individuals. Although reliabilities and other psychometric properties of within-person variability of accuracy measures might be stronger with different cognitive variables, or with additional sessions of measurement, the assessment of cognition in this project was fairly broad, and in most testing situations it may not be practical to administer more than three separate versions of each test.

In conclusion, the results of this study suggest that it is important to distinguish different types of within-person variability, as variability in accuracy across three sessions appears to have different properties from variability in RT across multiple trials within a single session. In light of its weak psychometric properties, within-person across-session variability in accuracy of cognitive performance does not appear to provide unique information about an individual's cognitive (and possibly clinical) status, and thus it may not be a useful individual difference characteristic.

Acknowledgments

Funding: The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was supported by Award Number R37AG024270 from the National Institute on Aging.

Appendix

Description of Reference Variables and Sources of Tasks

Variable	Description	Source
Wechsler Adult Intelligence Scale vocabulary	Provide definitions of words	Wechsler (1997a)
Picture vocabulary	Name the pictured object	Woodcock and Johnson (1990)
Antonym vocabulary	Select the best antonym of the target word	Salthouse (1993b)
Synonym vocabulary	Select the best synonym of the target word	Salthouse (1993b)
Matrix reasoning	Determine which pattern best completes the missing cell in a matrix	Raven (1962)
Shipley abstraction	Determine the words or numbers that are the best continuation of a sequence	Zachary (1986)
Letter sets	Identify which of five groups of letters is different from the others	Ekstrom French, Harman, and Dermen (1976)
Spatial relations	Determine the correspondence between a three-dimensional figure and alternative two-dimensional figures	Bennett, Seashore, and Wesman (1997)
Paper folding	Determine the pattern of holes that would result from a sequence of folds and a punch through folded paper	Ekstrom et al. (1976)
Form boards	Determine which combinations of shapes are needed to fill a larger shape	Ekstrom et al. (1976)
Logical memory Free recall	Number of idea units recalled across three stories Number of words recalled across Trials 1 to 4 of a word list	Wechsler (1997b)Wechsler (1997b)
Paired associates	Number of response terms recalled when presented with a stimulus term	Salthouse, Fristoe, and Rhee (1996)
Digit symbol	Use a code table to write the correct symbol below each digit	Wechsler (1997a)
Letter comparison	Same/different comparison of pairs of letter strings	Salthouse & Babcock (1991)
Pattern comparison	Same/different comparison of pairs of line patterns	Salthouse & Babcock (1991)

References

Adams ZW, Roberts WM, Milich R, Fillmore MT. Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? Psychological Assessment. 2011; 23:427–436. [PubMed: 21443365]

Bennett, GK.; Seashore, HG.; Wesman, AG. Differential aptitude test. San Antonio, TX: Psychological Corporation; 1997.

Bielak AAM, Hultsch DF, Strauss E, MacDonald SWS, Hunter MA. Intraindividual variability is related to cognitive change in older adults: Evidence for within-person coupling. Psychology and Aging. 2010; 25:575–586. [PubMed: 20853965]

Bleiberg J, Garmoe WS, Halpern EL, Reeves DL, Nadler JD. Consistency of within-day and across-day performance after mild brain injury. Neuropsychiatry, Neuropsychology, and Behavioral Neurology. 1997; 10:247–253.

Burton CL, Strauss E, Hultsch DF, Moll A, Hunter MA. Intraindividual variability as a marker of neurological dysfunction: A comparison of Alzheimer's disease and Parkinson's disease. Journal of Clinical and Experimental Neuropsychology. 2006; 28:67–83. [PubMed: 16448976]

Christensen H, Deary KBG, Anstey KJ, Parslow RA, Sachdev P, Jorm AF. Within-occasion intraindividual variability and preclinical diagnostic status: Is intraindividual variability an indicator of mild cognitive impairment? Neuropsychology. 2005; 19:309–317. [PubMed: 15910117]

Deary IJ, Der G. Reaction time explains IQ's association with death. Psychological Science. 2005; 16:64–69. [PubMed: 15660853]

Der G, Deary IJ. Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey. Psychology and Aging. 2006; 21:62–73. [PubMed: 16594792]

Dixon RA, Garrett DD, Lentz TL, MacDonald SWS, Strauss E, Hultsch DF. Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. Neuropsychology. 2007; 21:381–399. [PubMed: 17484601]

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

Ferrer E, McArdle JJ. Longitudinal modeling of developmental changes in psychological research. Current Directions in Psychological Science. 2010; 19:149–154.

Hultsch DF, MacDonald SWS, Dixon RA. Variability in reaction time performance of younger and older adults. Journal of Gerontology: Psychological Sciences. 2002; 57B:P101–P115.

Hultsch DF, MacDonald SWS, Hunter MA, Levy-Bencheton J, Strauss E. Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. Neuropsychology. 2000; 14:588–598. [PubMed: 11055261]

Hultsch, DF.; Strauss, E.; Hunter, MA.; MacDonald, SWS. Intraindividual variability, cognition, and aging. In: Craik, FIM.; Salthouse, TA., editors. The handbook of aging and cognition. 3rd. New York, NY: Psychology Press; 2008. p. 491p. 556

Li SC, Aggen SH, Nesselroade JR, Baltes PB. Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur successful aging studies. Gerontology. 2001; 47:100–116. [PubMed: 11287736]

Lovden M, Li SC, Shing YL, Lindenberger U. Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin aging study. Neuropsychologia. 2007; 45:2827–2838. [PubMed: 17575988]

MacDonald SWS, Hultsch DF, Dixon RA. Performance variability is related to change in cognition: Evidence from the Victoria longitudinal study. Psychology and Aging. 2003; 18:510–523. [PubMed: 14518812]

MacDonald SWS, Hultsch DF, Dixon RA. Predicting impending death: Inconsistency in speed is a selective and early marker. Psychology and Aging. 2008; 23:595–607. [PubMed: 18808249]

Nesselroade JR, Salthouse TA. Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 2004; 59(2):P49–P55.

Rabbitt P, Osman P, Moore B, Stollery B. There are stable individual differences in performance variability, both from moment to moment and from day to day. Quarterly Journal of Experimental Psychology. 2001; 54A:981–1003. [PubMed: 11765745]

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

Salthouse TA. Implications of within-person variability in cognitive and neuropsychological functioning on the interpretation of change. Neuropsychology. 2007; 21:401–411. [PubMed: 17605573]

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

Salthouse TA, Berish DE. Correlates of within-person (across-occasion) variability in reaction time. Neuro-psychology. 2005; 19:77–87.

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

Salthouse TA, Nesselroade JR. Dealing with short-term fluctuation in longitudinal research. Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 2010; 65:698–705.

Salthouse TA, Nesselroade JR, Berish DE. Short-term variability and the calibration of change. Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 2006; 61:P144–P151.

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

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

Saville CWN, Pawling R, Trullinger M, Daley D, Intri-ligator J, Klein C. On the stability of instability: Optimising the reliability of intra-subject variability of reaction times. Personality and Individual Differences. 2011; 51:148–153.

Schmiedek F, Lovden M, Lindenberger U. On the relation between mean reaction time and intraindividual reaction time variability. Psychology and Aging. 2009; 24:841–857. [PubMed: 20025400]

Shipley BA, Der G, Taylor MD, Deary IJ. Cognition and all-cause mortality across the entire adult age range: Health and lifestyle survey. Psychosomatic Medicine. 2006; 68:17–24. [PubMed: 16449407]

Strauss E, Bielak AAM, Bunce D, Hunter MA, Hultsch DF. Within-person variability in response speed as an indicator of cognitive impairment in older adults. Aging, Neuropsychology, and Cognition. 2007; 14:608–630.

Stuss DT, Murphy KJ, Binns MA, Alexander MP. Staying on the job: The frontal lobes control individual performance variability. Brain. 2003; 126:2363–2380. [PubMed: 12876148]

Wechsler, D. Wechsler Adult Intelligence Scale–third edition. San Antonio, TX: Psychological Corporation; 1997a.

Wechsler, D. Wechsler Memory Scale–third edition. San Antonio, TX: Psychological Corporation; 1997b.

Williams BR, Hultsch DF, Strauss EH, Hunter MA, Tannock R. Inconsistency in reaction time across the life span. Neuropsychology. 2005; 19:88–96. [PubMed: 15656766]

Woodcock, RW.; Johnson, MB. Woodcock-Johnson psycho-educational battery–revised. Allen, TX: DLM; 1990.

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

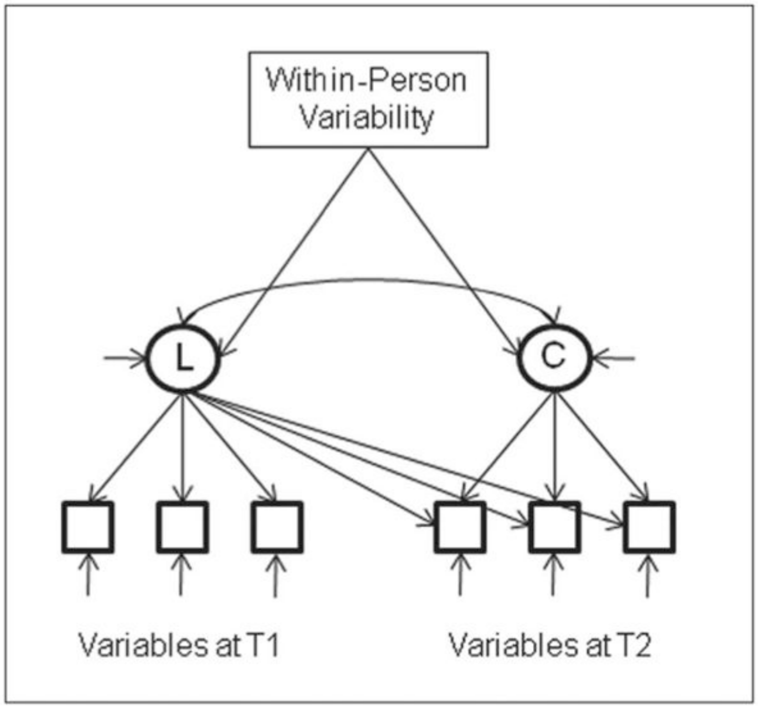

Figure 1.
Illustration of Latent Change Model Used to Estimate Relations Between Within-Person Variability and Longitudinal Change in Cognitive Ability
Note. L = latent level; C = latent change.

Table 1

Sample Characteristics

	18-59 years	60-95 years
Total sample		
N	1,067	658
Age	41.0 (13.5)	71.7 (8.1)
Proportion of females	0.68	0.59
Health	2.2 (0.9)	2.4 (0.9)
Years of education	15.4 (2.5)	16.1 (2.9)
MMSE	28.7 (1.6)	28.1 (2.0)
Scaled scores		
Vocabulary	12.3 (3.2)	13.0 (2.6)
Digit symbol	11.1 (2.8)	11.4 (2.8)
Logical memory	11.5 (2.9)	12.1 (2.8)
Word recall	11.8 (3.4)	12.1 (3.5)
Longitudinal sample at Time 1		
N	335	244
Age	42.8 (13.4)	71.7 (7.5)
Proportion of females	0.67	0.59
Health	2.3 (0.9)	2.4 (0.9)
Years of education	15.4 (2.5)	16.1 (2.8)
MMSE	28.8 (1.5)	28.3 (1.8)
Scaled scores		
Vocabulary	12.5 (2.8)	13.5 (2.6)
Digit symbol	11.4 (2.8)	11.8 (2.6)
Logical memory	11.6 (2.8)	12.6 (2.7)
Word recall	12.2 (3.6)	12.8 (3.3)
Time 1 to Time 2 interval (years)	2.3 (0.7)	2.4 (0.6)

Note. Health was a self-rating on a scale from 1 = *excellent* to 5 = *poor*. MMSE = Mini Mental Status Exam (Folstein, Folstein, & McHugn, 1975). Scaled scores are age-adjusted scores that have means of 10 and standard deviations of 3 in the normative standardization samples (Wechsler, 1997a, 1997b).

Table 2

Means (and Standard Deviations) of Within-Person Variability at Time 1

Total sample	**Within-person variability (*SD*)** 18-59 years	60-95 years	*d*
Variable			
Vocabulary	.22 (.19)	.22 (.22)	0
Picture vocabulary	.21 (.16)	.21 (.16)	0
Synonym vocabulary	.28 (.24)	.26 (.22)	−.09
Antonym vocabulary	.27 (.22)	.29 (.24)	.09
Matrix reasoning	.25 (.20)	.27 (.22)	.10
Shipley abstraction	.33 (.28)	.34 (.25)	.04
Letter sets	.27 (.23)	.30 (.26)	.12
Spatial relations	.27 (.20)	.28 (.21)	.05
Paper folding	.23 (.18)	.24 (.19)	.05
Form boards	.25 (.21)	.21 (.19)	−.20
Word recall	.22 (.17)	.24 (.19)	.11
Paired associates	.52 (.25)	.50 (.23)	−.08
Logical memory	.23 (.18)	.24 (.21)	.05
Digit symbol	.18 (.18)	.16 (.16)	−.12
Pattern comparison	.22 (.20)	.22 (.20)	0
Letter comparison	.22 (.22)	.22 (.22)	0

	Time 1			Time 2		
	Within-person variability (*SD*)			Within-person variability (*SD*)		
Longitudinal sample	18-59 years	60-95 years	*d*	18-59 years	60-95 years	*d*
Variable						
Vocabulary	.20 (.16)	.20 (.18)	0	.19 (.19)	.22 (.22)	.15
Picture vocabulary	.20 (.16)	.22 (.16)	.13	.18 (.15)	.20 (.17)	.13
Synonym vocabulary	.27 (.22)	.25 (.21)	−.09	.29 (.23)	.28 (.20)	−.05
Antonym vocabulary	.28 (.22)	.27 (.20)	−.05	.26 (.19)	.28 (.22)	.10
Matrix reasoning	.25 (.20)	.29 (.20)	.20	.26 (.22)	.26 (.23)	0

	Within-person variability (*SD*)					
Total sample	**18-59 years**	**60-95 years**	***d***			
Shipley abstraction	.34 (.28)	.33 (.23)	−.04	.21 (.17)	.25 (.17)	.24
Letter sets	.27 (.28)	.30 (.23)	.12	.22 (.18)	.30 (.25)	.38
Spatial relations	.27 (.22)	.26 (.19)	−.05	.25 (.20)	.27 (.21)	.10
Paper folding	.21 (.20)	.24 (.17)	.16	.23 (.20)	.23 (.17)	0
Form boards	.23 (.17)	.19 (.16)	−.24	.24 (.17)	.22 (.18)	−.11
Word recall	.22 (.17)	.22 (.17)	0	.20 (.17)	.24 (.19)	.22
Paired associates	.53 (.26)	.50 (.22)	−.12	.53 (.26)	.50 (.22)	−.12
Logical memory	.24 (.19)	.24 (.19)	0	.22 (.19)	.22 (.18)	0
Digit symbol	.17 (.16)	.16 (.17)	−.06	.21 (.23)	.20 (.22)	−.04
Pattern comparison	.21 (.19)	.22 (.19)	.05	.23 (.23)	.18 (.15)	−.25
Letter comparison	.20 (.19)	.23 (.22)	.15	.29 (.32)	.24 (.22)	−.18

* $p < .01$.

Table 3

Stability (Time 1 to Time 2) Correlations for Session 1 Scores and for Within-Person Across-Session Variability (*SD*)

	Session 1 score		Within-person variability (*SD*)	
Variable	**18-59 Years**	**60-95 Years**	**18-59 Years**	**60-95 Years**
Vocabulary	.86*	.74*	.14*	.12
Picture vocabulary	.89*	.83*	.22*	.14
Synonym vocabulary	.85*	.72*	.20*	.18*
Antonym vocabulary	.76*	.57*	.11	.10
Matrix reasoning	.74*	.66*	.04	.08
Shipley abstraction	.87*	.83*	−.02	−.05
Letter sets	.70*	.61*	.07	.04
Spatial relations	.86*	.79*	.04	.17*
Paper folding	.72*	.65*	.07	−.03
Form boards	.72*	.69*	.02	.19*
Word recall	.69*	.66*	.11	.24*
Paired associates	.67*	.65*	.42*	.18*
Logical memory	.65*	.65*	.05	.08
Digit symbol	.77*	.74*	.08	−.06
Pattern comparison	.66*	.71*	.02	−.02
Letter comparison	.70*	.71*	.07	−.06

* $p < .01$.