



Published in final edited form as:

Psychol Aging. 2005 September ; 20(3): 390–401. doi:10.1037/0882-7974.20.3.390.

Intraindividual Variability May Not Always Indicate Vulnerability in Elders' Cognitive Performance

Jason C. Allaire and

Department of Psychology, North Carolina State University

Michael Marsiske

Department of Clinical and Health Psychology, University of Florida

Abstract

This study examined consistency of performance, or intraindividual variability, in older adults' performance on 3 measures of cognitive functioning: inductive reasoning, memory, and perceptual speed. Theoretical speculation has suggested that such intraindividual variability may signal underlying vulnerability or neurologic compromise. Thirty-six participants aged 60 and older completed self-administered cognitive assessments twice a day for 60 consecutive days. Intraindividual variability was not strongly correlated among the 3 cognitive measures, but, over the course of the study, intraindividual variability was strongly intercorrelated within a task. Higher average performance on a measure was associated with greater performance variability, and follow-up analyses revealed that a higher level of intraindividual variability is positively associated with the magnitude of a person's practice-related gain on a particular measure. The authors argue that both adaptive (practice-related) and maladaptive (inconsistency-related) intraindividual variability may exist within the same individuals over time.

Keywords

inconsistency; older adults; cognitive aging; age effects; intraindividual variability in cognitive performance

Short-term intraindividual variability represents transient within-person fluctuations in behavioral performance. Some authors have argued that such variability represents the steady-state "hum" of psychological constructs (Ford, 1987; Nesselrode, 1991). Li, Huxhold, and Schmiedek (2004) have argued that such short-term intraindividual variability must be distinguished from longer term and more enduring behavioral changes, which more typically would be referred to as development or change (Baltes, Reese, & Nesselrode, 1977; Nesselrode, 1991). The goal of the present study was to explore short-term intraindividual variability in cognition in a sample of community-dwelling older adults and to further understand the association of such intraindividual variability with individual differences in cognitive performance level. The current study also examined whether within-person variability itself remained a stable (i.e., trait-like) attribute of individuals over a 2-month period of repeated measurement.

Correspondence concerning this article should be addressed to Jason C. Allaire, Department of Psychology, North Carolina State University, Campus Box 7801, Raleigh, NC 27695-7801 or Michael Marsiske, Department of Clinical and Health Psychology, University of Florida, PO Box 100165, 101 South Newell Drive, Room 3151, Gainesville, FL 32610-0165. jason_allaire@ncsu.edu or marsiske@ufl.edu.

This study extends previous research in several ways. First, we examined intraindividual variability in performance accuracy on a set of primarily fluid or mechanic abilities (inductive reasoning, list memory, perceptual speed). Although such measures have been widely studied in the cognitive aging literature (Ball et al., 2002; Singer, Verhaeghen, Ghisletta, Lindenberger, & Baltes, 2003), much of the gerontological research on intraindividual variability has focused on reaction time data (Anstey, 1999; Hultsch, MacDonald, & Dixon, 2002; MacDonald, Hultsch, & Dixon, 2003; Rabbitt, Osman, Moore, & Stollery, 2001; Strauss, MacDonald, Hunter, Moll, & Hultsch, 2002). Second, we included a substantially larger number of assessment occasions (120 occasions; twice-daily assessments for 60 days) than has typically been possible. Third, given the nature of the measures used and the number of assessments, we expected to see substantial retest-related improvements in performance over the course of study. Rather than treat such retest related gains as nuisance, we chose to examine whether there was an association between improvement over occasions and intraindividual variability. There are indications from research earlier in the life span as well as from one previous gerontological investigation that intraindividual variability may, in some circumstances, be a reflection of learning and strategy modifications (Li, Aggen, Nesselroade, & Baltes, 2001; Li et al., 2004; Rittle Johnson & Siegler, 1999; Siegler, 1994).

Methodological Issues in Studies of Intraindividual Variability

It is difficult to abstract common findings across previous studies. Studies have varied in the types of participants assessed. In addition, across investigations, researchers have examined a heterogeneity of tasks, typically using a variety of simple and complex reaction time measures.¹ In this study, we wanted to examine more closely those domains of cognition (i.e., memory, reasoning, speed) that have typically been included in the major longitudinal (see Schaie & Hofer, 2001, for a review) and intervention (Ball et al., 2002) studies of intellectual–cognitive development. Specifically, our goal was to examine within-person variability in domains for which the adult developmental trajectories of means and covariances are already very well understood.

A second source of inconsistency in extant studies has been the temporal resolution of the intraindividual variability measure used. Some studies have examined momentary variation, as assessed by trial-to-trial latency assessed over multiple days or longer. Other studies have examined different interassessment intervals, and the relationship between these different temporal resolutions is not known. Assessment frequencies have ranged from multiple trials at one occasion (Anstey, 1999); to multiple trials at several occasions over a few days (West, Murphy, Armilio, Craik, & Stuss, 2002), weeks, or months (Fuentes, Hunter, Strauss, & Hultsch, 2001; Hultsch et al., 2002; MacDonald et al., 2003; Rabbitt et al., 2001); to single assessments spaced over a number of months (Li et al., 2001; Rabbitt et al., 2001). Thus, a comparison of the magnitude of intraindividual variability across studies is difficult. As pointed out by Martin and Hofer (2004), intraindividual variabilities in performance over differing temporal frames (i.e., high resolution vs. low resolution) are likely to be different but related phenomena, but this remains an empirical question. In the present investigation,

¹There are good justifications for examining reaction times. Reaction time measures have large potential ranges, which means that they are more sensitive to individual differences than traditional cognitive test accuracy scores, which will typically allow for much less variability, bounded by the number of items that a test can include. In addition, a key underlying conception is that variability is a reflection of underlying neural health or cognitive efficiency (Li & Lindenberger, 1999), and measures of basic information processes are thought to be more proximal indicators of such basic functioning. Third, typical reaction time tasks include many trials, which allow for the efficient collection of multiple behavior samples needed to conduct a within-person variability study. Fourth, some reaction time tasks are relatively less sensitive to retest effects or relatively quickly evince performance asymptotes in participants. One intent in this article, however, was to examine whether the more typically used accuracy or correctness measures of variability used in the cognitive aging literature behaved similarly to latency measures.

we focused on intraindividual variability in a daily resolution context because of the larger goals of our project, which sought to investigate the coupling of daily variability in activity, physiology, sensorimotor parameters, and affect with cognition.

Consistency of Intraindividual Variability Over Measures

Empirical findings suggest that within-person variability does not represent purely idiosyncratic, measure-specific error but rather is correlated across multiple measures (i.e., convergent validity). For instance, Hultsch et al. (Burton, Hultsch, Strauss, & Hunter, 2002; Fuentes et al., 2001; Hultsch et al., 2002; Hultsch, MacDonald, Hunter, Levy Bencheton, & Strauss, 2000; Mac-Donald et al., 2003; Strauss et al., 2002) have consistently reported that intraindividual variability across different measures of reaction time (i.e., simple, complex, visual search, word recognition, lexical, and semantic decision tasks) is strongly correlated not only within a session containing multiple trials but across multiple testing sessions as well. Similarly, Li et al. (2001) found evidence of positive associations among estimates of intraindividual variability for sensorimotor functioning (i.e., walking). Covariation of the intraindividual variability estimates among some of the cognitive variables as well as between some of the sensorimotor and cognitive variables was reported.

Consistency of Intraindividual Variability Over Occasions

Prior studies have focused on investigating whether intraindividual variability is a consistent property or trait of the individual. In general, most studies have reported findings suggesting that variability remains fairly stable within a testing session consisting of multiple trials as well as across testing occasions (Fuentes et al., 2001; Hultsch et al., 2000; Nesselroade & Salthouse, 2004; Rabbitt et al., 2001). However, Li et al. (2001) have reported findings that contradict, to an extent, these conclusions. These authors found that, although sensorimotor (i.e., walking performance) intraindividual variability in the first half of the study was strongly correlated with variability in second half of the study, the same was not true for the measures of memory. That is, individuals who exhibited more intraindividual variability in memory performance during the first 12 occasions were not necessarily the same individuals who exhibited more variability in the final 13 occasions.

Variability Signals Vulnerability

In addition to examining patterns of covariation, previous studies have focused on the association between intraindividual variability and level of cognitive performance. Guided by the hypothesis that increased intraindividual variability in cognitive functioning is indicative of compromised neurobiological functioning (Bruhn & Parsons, 1977; Hendrickson, 1982; Li & Lindenberger, 1999), most cognitive aging researchers have considered intraindividual variability as a reflection of a compromised cognitive system. Indeed, prior work has found that greater intraindividual variability in performance on a specific cognitive measure is associated with poorer performance on that measure (Anstey, 1999; Fuentes et al., 2001; Hultsch et al., 2000; Salthouse, 1993), poorer performance on different cognitive tests assessing multiple domains of cognition (Hultsch et al., 2002; Nesselroade & Salthouse, 2004; Rabbitt et al., 2001), and longitudinal decline in a number of basic cognitive abilities (MacDonald et al., 2003). Moreover, studies have found increased intraindividual variability in the presence of different clinical conditions such as dementia (Hultsch et al., 2000), chronic fatigue syndrome (Fuentes et al., 2001), and brain injury (Stuss, Pogue, Buckle, & Bondar, 1994; West et al., 2002). This association does not appear to be due to ceiling or floor effects in the measure studied.

One exception to the general finding that intraindividual variability signals vulnerability comes from the work of Li et al. (2001), who found mixed results regarding the relationship

between mean and intraindividual variability. Specifically, they reported that the association between intraindividual variability and mean level of performance was in the expected negative direction for two measures of story recall (i.e., gist recall and elaboration). In contrast, positive associations between level and intraindividual variability were found for forward and backward digit span and two measures of spatial memory, although the relationships were only significant for the spatial memory tasks; higher overall performance was related to greater intraindividual variability on these measures. As mentioned, these authors report that variability in early occasions was not strongly related to variability in later occasions, possibly suggesting that practice-related reorganizations of memory performance over time also altered the nature of the underlying performance intraindividual variability.

Li et al.'s findings differ from the extant aging research on cognitive intraindividual variability; however, they do mesh with the prevailing conceptualizations of intraindividual variability in the child development literature (van Geert & van Dijk, 2002). Here intraindividual variability in domains such as motor functioning (Piek, 2002), mother-child interactions (de Weerth & van Geert, 2002), emotion (de Weerth, van Geert, & Hoijsink, 1999), language (van Dijk, de Goede, Ruhland, & van Geert, 2001), and cognition (Rittle Johnson & Siegler, 1999; Siegler, 1994; Siegler & Lemaire, 1997) is considered a potential marker of developmental change and improvement within these domains. Intraindividual variability might represent the testing of new behaviors or strategies that lead to more developed behaviors and modes of thinking (Siegler, 1994; Siegler & Lemaire, 1997; van Geert & van Dijk, 2002). Guided by this literature, Li et al. characterized the positive association between intraindividual variability and mean as adaptive intraindividual variability, which may stand in contrast to the more typical maladaptive intraindividual variability reported in most cognitive aging studies to date (e.g., Li et al., 2004).

Consequently, in the present investigation, we chose to examine the magnitude of individual performance gain as an independent correlate of intraindividual variability. In addition, in a long run of trials such as ours (120 occasions), it was also possible to investigate whether there were systematic changes in intraindividual variability from early to late practice occasions. We hypothesized that, in fact, some of the observed intraindividual variability might reflect underlying learning- and practice-related processes. Thus, the study addressed the following questions: To what extent was intraindividual variability homogeneous across multiple cognitive abilities (i.e., inductive reasoning, memory, and perceptual speed)? Is intraindividual variability on measures of these abilities a stable individual difference characteristic? What is the pattern of intraindividual variability and mean performance, and does it remain stable over the course of the study?

Method

Sample

Data were taken from 36 participants who completed an extensive daily assessment protocol. The sample consisted of 10 men and 26 women, all of whom were community-dwelling elders recruited from local senior centers. Participants ranged in age from 60 to 87 years ($M = 74$ years, $SD = 5.51$). The sample of older adults was highly educated (average years of education = 16, $SD = 2.98$ years, range = 12–22 years) and had an average household income of \$28,000 ($SD = \$6,000$, range = \$12,000–50,000+). In addition, 34 of the participants identified themselves as White and 2 reported their racial status as African American. Participants rated their health and sensory functioning compared with other same-aged individuals on a 6-point Likert-type scale ranging from 1 (*very good*) to 6 (*very poor*). Average ratings on these items indicated that the participants believed their functioning in these domains to be between good and moderately good. Specifically, the

mean ratings were as follows: general physical health, 1.91 ($SD = 0.81$); vision, 2.61 ($SD = 0.87$); and hearing, 2.83 ($SD = 1.36$). Frequency of experiencing positive and negative emotions was also assessed (1 = *not at all*, 5 = *very much*). Ratings indicated a high level of positive emotion ($M = 3.49$, $SD = 0.68$, range = 2.3–4.90) and a low level of negative emotion ($M = 1.43$, $SD = 0.49$, range = 1.00–2.80).

Procedure

The design of the current study is considered an extension of the multivariate, replicated, single-subject, repeated measures design (Nesselroade & Ford, 1987; Nesselroade & Jones, 1991). Participants were asked to complete a daily mental exercise workbook twice a day, once in the morning and once at night, for 60 consecutive days for a total of 120 occasions of measurement. Additional information regarding the content and development of the cognitive measures included in the daily mental exercise workbook is provided in the *Measures* section.

Before beginning the daily assessment, each participant attended an orientation session in groups of 2 to 5 individuals where they received instructions on how to self-administer the measures in the workbook. A trained proctor guided the participants through an example workbook, thoroughly explaining the instructions for each of the measures, including how to correctly use an electronic timer, which was needed for the cognitive tests. Once a particular measure was described and examples were worked through together, participants practiced independently self-administering the measures under the supervision of project staff. Corrective feedback was provided, and additional instructions as well as practice were allowed until all participants in a session felt comfortable with the procedures.

Participants were given four guidelines that they were asked to adhere to over the course of the daily assessment: (a) Two workbooks must be completed every day, one in the morning (within 2 hr of waking) and one at night (within 2 hr of going to bed) for 60 consecutive days; (b) workbooks were to be completed without receiving outside assistance; (c) in the event that a workbook was skipped it was to be left blank instead of making it up later; (4) completed workbooks were to be mailed back to the project office using envelopes with prepaid postage at the end of each week.

After the training was completed and these guidelines were reviewed, participants were provided with the testing materials needed for the duration of the daily assessment, which included 120 predated and presorted daily mental exercise workbooks, a digital timer, and an automatic blood pressure monitor. Workbooks were clearly marked with the corresponding date as well as the time of day (i.e., morning or evening) it was to be completed. Although brief instructions for each measure were included in each workbook, participants also received a detailed daily mental exercise instructional manual, which provided additional instructions for each measure as well as examples. Participants were given day and evening contact numbers for study personnel in the event that additional questions or problems arose over the course of the daily assessment.

Although the self-administration is a unique aspect of the current study, it does allow one to speculate on the fidelity of the data that were collected. Consequently, a number of quality assurance strategies were used in the attempt to minimize deviations from protocol. First, participants were asked to verify that they did not receive assistance in completing the workbook by signing a declaration of noncollaboration on the front page of each workbook, which was signed by all of the participants on 100% of the 120 occasions. Second, weekly returned workbooks were reviewed for completeness, and participants were contacted if a pattern of mistakes or skipped items was found. Third, a message was stapled to the cover of two randomly selected workbooks each week instructing participants to call an answering

machine and leave a message that included their identification number. The answering machine stamped their message with the date and time, and those individuals who did not call in were contacted by phone to determine whether they were facing any difficulties with completing their workbooks at the appropriate time. In addition to these steps, preliminary analysis examined the association between performance on the supervised pretest measures (Letter Series Test, Number Comparison Test, and List Memory) and the mean of each participant's first five unsupervised testing occasions. Results indicated that performance was significantly correlated ($r_s = .52-.75$), and no significant mean differences were found between the pretest and the mean of the first five occasions.

Measures

As mentioned earlier, participants in the current study were assessed twice a day for 60 consecutive days by completing what are referred to as daily mental exercise workbooks. These workbooks contained three cognitive measures assessing inductive reasoning as assessed by the Letter Series Test (Thurstone, 1962), perceptual speed as assessed by the Number Comparison Test (Ekstrom, French, Harman, & Derman, 1976), and list memory as assessed by Rey's Auditory Verbal Learning Task (AVLT; Rey, 1941). Table 1 contains additional information regarding the cognitive measures. In addition to these tests, the workbooks included measures from noncognitive domains such as activity engagement, perceived control, positive and negative affect, blood pressure and pulse, and reading vision.

Alternate test versions—To minimize the effect of test familiarity, we used 14 alternate forms of the three cognitive measures. Therefore, each of the 14 tests used in a given week were different versions of each measure, and participants were not administered the same version twice within any given week. The order in which alternate forms were encountered in the workbook was counterbalanced across participants. It is important to note that alternate version is covaried out when creating our index of intraindividual variability to control for subthreshold form effects.

The items in each of the 14 alternate versions of the Letter Series Test were equated with respect to (a) the difficulty of the pattern in the string of letters and (b) the difficulty of the distractors included in answer choices. To construct the alternate forms, the parent measure was first parsed into the specific rules and strategies that defined each item. Then, by simple alphabetical shifts to each letter series, new items that followed the same rules as the parent items were generated. For instance, an alternate version for the item “a a b b c c d ___” would be “b b c c d d e ___” and a second alternate version would be “c c d d e e f ___.” In addition, possible answer choices were also changed, which was accomplished by shifting each of the letters in the answer choices by one place in the alphabet for each of the 14 alternative versions. Participants were given 4 min to complete as many items as they could.

In the parent measure of the Number Comparison Test, 28 of the numerical string pairs differed by a single number (e.g., 376, 386). The difficulty of these items is determined by two factors: (a) the length of each numerical string (i.e., 3 to 13 numbers) and (b) the serial position of the discrepant number within the string. Fourteen alternate versions of the Number Comparison Test were developed by randomly changing the numbers in each numerical string while retaining the difficulty of each item. That is, the length of each numerical string and the position of the aberrant number within the string were retained across alternate versions while altering the numbers in each version. Participants were given 90 s to complete as many items as possible.

Thirteen of the alternate word lists for the AVLT list memory task used in the current study were taken from previous published parallel forms of the AVLT (Schmidt, 1996), and an additional alternative word list was created for the purposes of the current study. The

selection of the words comprising this additional list was based on their similarity to the preexisting lists with respect to their frequency of appearance in printed text. Analysis indicated that the 15 words that made up the list created for the current study did not, on average, differ in frequency from the words in 13 preestablished lists, $F(13, 209) = 1.05, p > .05$. Participants were instructed to study the list of words for 1 min and then turn the page and were allowed an additional minute to write as many words as they could remember.

Calculating intraindividual variability—To quantify within-person variability, an index representing intraindividual variability was created for each participant separately for each of the three cognitive tests.² This index was calculated in four steps. First, a regression equation was estimated for each cognitive test, separately for each participant, in which the dependent variable was cognitive performance at each occasion. The independent variables were the linear and quadratic effects of occasion, which were included to control for intraindividual differences in learning. The intent was to detrend the data before computation of the intraindividual variability index; failure to do so would inflate intraindividual variability due to practice-related improvements over occasions. In addition to the time trends, dummy variables representing the alternate forms were included to control for possible form differences due to difficulty. This technique also covaried out diurnal variations from these data. Second, the difference between each participant's predicted value (based on the regression line) and their obtained value at each occasion of measurement (i.e., residual) was calculated and squared to remove negative values. Third, these squared residuals were then summed over all occasions for a participant, and this sum was divided by the number of available occasions (i.e., mean squared residual). Fourth, the square root of this mean squared residual was obtained for each participant on each of the three cognitive variables; this put the metric of the index back in the original scale of the instrument. We labeled the resulting term the intraindividual residual index (IRI), and it represented the average amount of intraindividual variability in participants' performance around their personal best fitting regression line. Figure 1 provides a graphic representation of the estimated regression line fit to the Letter Series Test raw data of 2 participants differing in mean performance (i.e., 1 *SD* above and below the mean).

A second index representing participants' mean performance over the entire course of the daily assessment was created for each measure. This intraindividual mean was calculated by averaging within each participant across the 120 occasions of measurement. It is important to note that both the IRIs and mean scores were estimated using only the available data for each participant. With respect to the amount of missing data, on average participants completed the contents of 115 ($SD = 7.28$) of 120 workbooks. Of the possible 4,320 data points (36 participants \times 120 occasions), only 166 were missing.

To examine intraindividual variability over the course of the daily assessments, the 120 measurements were divided into four time-ordered blocks, with an equal number of occasions in each block (i.e., 1–30, 31–60, 61–90, 91–120). The decision to create four blocks was somewhat arbitrary; however, two blocks would have provided too little precision to detect smaller transitions in overall intraindividual variability, whereas more than four blocks would reduce the number of occasions contributing to block-specific intraindividual variability estimates, thereby reducing the reliability of those estimates.

²In data with learning or practice trends, the simple computation of within-subjects variability as a within-person standard deviation is not appropriate. Standard deviations are computed around a person's overall mean across all occasions; for participants with high levels of improvement, the difference between early and late performance points and their overall mean will tend to be greater than for participants with minimal learning, leading to overestimates of variability for those who learn the most. Thus, data need to be detrended for learning effects. In this study, we chose to detrend by covarying on linear and quadratic occasion effects.

Using the approach described previously, intraindividual variability estimates were calculated within each of the four blocks for the three cognitive tests (see Appendix).

Sample attrition—Demands on participants exceeded those of typical studies of cognitive aging, given that participants had to engage in self-administered testing twice a day for 2 months. Thus, as in physical exercise trials with older adults, there were individual differences in compliance (Orsega Smith, Payne, & Godbey, 2003); therefore, we expected a positively selected subset of the original sample of recruited individuals at the conclusion of the study. Correspondingly, 50 older adults had agreed to participate, and 14 participants dropped out before the completion of the study. Reasons for dropout fell into one of two categories: Participants voluntarily terminated their participation, or they were dropped from the study because of failure to comply with the repeated assessment protocol (i.e., completed less than 75% of the daily assessments). Comparisons between those who completed the study and those who did not are provided in Table 2. Individuals who did not complete the study were, on average, younger and less educated, reported a lower level of income, and performed more poorly on the Letter Series Test than those who completed the study. In addition, some daily performance data were available for 9 of the 14 participants who dropped out of the study. Analysis examining intraindividual variability using any available data within the first 30 occasions revealed no significant differences between the dropouts and completers for the Letter Series and Number Comparison Tests. However, there was a significant difference in intraindividual variability for the AVLT list memory task, $t(43) = -8.81, p < .05$; the average intraindividual variability index of the dropped out participants ($M = 1.76, SD = 0.35$) was significantly lower than that of participants who completed the study ($M = 6.43, SD = 1.23$).

Results

The purpose of the current study was to examine intraindividual variability in older adults' cognitive functioning as assessed by three cognitive ability measures. Specifically, three research questions were addressed in this investigation, the results of which are presented in the following section. First, to what extent is intraindividual variability associated across cognitive measures? Second, are individual differences in intraindividual variability consistent over the course of practice? Third, what is the relationship between mean performance and intraindividual variability? The last two questions were addressed across the 120 occasions as well as within the four blocks of 30 occasions.

Association Among Intraindividual Variability Estimates

The correlations among the intraindividual variability estimates for the three cognitive measures were not uniform. Specifically, the relationship between variability in the Letter Series Test and variability in the Number Comparison Test was negative but failed to reach significance ($r = -.22$). Similarly, the relationship between variability for list memory and the Number Comparison Test was negative and nonsignificant ($r = -.30$). However, variability in Letter Series Test performance was significantly and positively related to variability in the AVLT list memory task ($r = .39$), indicating that more intraindividual variability in Letter Series Test performance was associated with greater intraindividual variability in performance on the list memory task. The relatively small sample size in this study was partially responsible for the findings of nonsignificance; the power to detect the two lower correlations in this study was 0.38 and 0.58, respectively³ (Cohen, 1988).

³Recognizing that some readers would prefer interval estimates of the correlations, we will provide a fuller table of correlations, with confidence intervals, on request.

Within-Person Stability of Intraindividual Variability

Analyses next examined the correlations among the intraindividual variability estimates calculated within each block, the results of which are presented in Table 3. A simplex pattern (Gorsuch, 1974) of relationships emerged for the Letter Series Test, particularly for Block 1. That is, intraindividual variability in Block 1 was positively and significantly correlated to Block 2 intraindividual variability but unrelated to variability in the more distal blocks (Blocks 3 and 4). The correlation between variability in Blocks 2 and 3 was positive but fell just short of statistical significance ($p < .10$); however, the association between variability in Blocks 2 and 4 as well as in Blocks 3 and 4 was positively and significantly related. Intraindividual variability was strongly and positively correlated over the four blocks for the Number Comparison Test. Similarly, the correlations over blocks for the AVLT list memory task were positive and significant.

Intraindividual Variability and Level of Performance

Analyses next examined the correlations between average performance and intraindividual variability over the 120 occasions. Contrary to expectations, the correlation between intraindividual variability and mean was positive for the three cognitive measures. Specifically, the correlation between intraindividual variability and the mean performance for the Number Comparison Test was significant ($r = .34, p < .05$), as was the correlation for list memory ($r = .43, p < .05$). The relationship between the variability and mean for the Letter Series Test was also positive ($r = .26, p > .05$) but failed to reach statistical significance. These positive correlations suggest that higher average performance was associated with greater intraindividual variability in performance.

Intraindividual variability and mean within block—Analyses next examined the correlations between intraindividual variability and mean performance within blocks, which are presented in Table 4. Starting with the Letter Series Test, variability in Block 1 and mean performance were significantly and positively correlated, indicating that more variability over the first 30 occasions was related to higher mean performance. Although not significant, the positive relationship between variability and mean performance was also present in the second block of occasions. However, in Blocks 3 and 4, which correspond to the second half of the daily assessment, the direction of the correlation between variability and overall mean performance was negative. Turning to the Number Comparison Test and list memory task, the within-block relationships between variability and overall level of performance were uniformly positive. Specifically, the correlations for the Number Comparison Test, although falling short of significance, were positive and moderate in magnitude. The correlations between within-block variability and mean performance for list memory were significant in Blocks 1 through 3 but failed to reach significance in Block 4.

One possibility for the negative relationships in Blocks 3 and 4 of the Letter Series Test might have been the presence of ceiling effects. Participants who mastered the test and who were consistently performing at the top of the distribution would have had less room to fluctuate, leading to a reduction in variability and a spurious negative correlation between mean performance and within-block variability. Seven participants were identified as performing at ceiling on the Letter Series Test and 5 performed at ceiling on the list memory task (i.e., scored incorrectly on no more than 1 item [e.g., 29 of 30] and remained at that level of performance for the remainder of the trials). Floor effects were not observed for any of the tests, and ceiling effects were not observed for the Number Comparison Test. Consequently, the correlation between mean performance and within-block variability for the Letter Series Test and list memory task were reexamined after eliminating participants who reached, and stayed, at ceiling. The within-block correlations, after dropping ceiling participants, are presented in Table 4. The magnitude of the mean–variability correlation for

the Letter Series Test increased in Blocks 1 and 2, whereas in Blocks 3 and 4 the once negative correlations were now positive. The correlations remained relatively unchanged for the AVLTL, with the exception of Block 4, which increased in magnitude.

Follow-Up Analysis

The pattern of results suggested that, for the current sample of older adults, mean performance and variability were positively interrelated. This is further supported by the negative relationship between mean and variability for the Letter Series Test when participants at ceiling were included and the positive association when these participants were excluded from the analysis. To elucidate the potential relationship between average performance and variability, additional analysis examined the temporal pattern (i.e., practice effects) for the Letter Series and Number Comparison Tests and the list memory task over the 120 occasions.

Using a mixed-effects modeling approach, a growth model was estimated for each test to determine the extent to which performance over the 120 occasions could be characterized by practice-related improvement. A separate growth model was estimated for each of the cognitive tests. In each model, the intercept was set to the first occasion, and the linear and quadratic time trends were recoded so that the unstandardized beta coefficients reflected the average amount of change over the 120 occasions. Performance on the Letter Series Test, after removing ceiling participants, exhibited strong linear [$B = 44.30, t(3308) = 13.63, p < .01$], and quadratic [$B = -23.15, t(3308) = -11.09, p < .01$], trends with practice-related improvement followed by a leveling off later in performance. The Number Comparison Test exhibited a linear pattern [$B = 12.86, t(4120) = 4.84, p < .01$], with some evidence of a quadratic trend [$B = -6.38, t(4120) = -2.57, p < .05$], with leveling off during the later half of the study. List Memory Test was best described by a linear trend [$B = 18.84, t(4121) = 4.50, p < .01$], with a steady increase seen over the course of the 120 occasions and a slight quadratic trend [$B = -6.96, t(4121) = -2.19, p < .05$]. The mixed models permitted the estimation of random effects for the linear and quadratic terms; these indicated that there were significant intraindividual differences in both linear and quadratic intraindividual change for all three cognitive measures studied. Figure 2 displays the average 120 occasion trajectories for the three cognitive tests. Based on these trends, it appears that the Letter Series Test has an asymptotic function. The Number Comparison and List Memory Tests are best characterized by linear practice-related improvement, and the significant quadratic effects were due to the power of the models to detect subtle changes in curvature. Related to the earlier results, and focusing specifically on the Letter Series Test, we noted that the relationship between intraindividual variability and mean was positive in those blocks of occasions in which the overall practice-related trend was linear and positive, as is the case for the first two blocks (i.e., Occasions 1–60) of the Letter Series Test and all four blocks for other two tests. Thus, the growth models led us to speculate that linear slope estimates for a participant ought to be positively related to his or her intraindividual variability. That is, in that segment of the practice or practice-related curve in which participants are gaining in performance, one might also expect to observe higher levels of intraindividual variability.

To test this speculation, we conducted additional post hoc analyses in which we examined the relationship between each participant's linear gain slope over time and their intraindividual variability. We first obtained the unstandardized linear slope estimate from the best fitting regression line that were estimated within each of the four blocks for each participant using the within-person regression approach used to calculate the IRI. The slope estimates were then correlated with the within-block variability estimates, as calculated earlier, for each of the cognitive tests. As can be seen in Table 5, the correlations between slope and variability for the Letter Series Test were positive and significant (with the exception of Block 3), such that the more linear growth participants exhibited within a

block, the more variability they exhibited. The correlations between slope and intraindividual variability for the Number Comparison Test were positive in Blocks 2 and 3. Variability and slope were also positively associated within each block for the List Memory Test (with the exception of Block 3).

Discussion

Theoretical interest and empirical research focusing on intraindividual variability has grown in the field of cognitive aging. The present study represents an attempt to contribute several ideas to this line of inquiry. The first research objective of this study was to determine the extent to which intraindividual variability in performance was associated across multiple cognitive abilities. That is, if a person evinces high intraindividual variability in one domain of cognition, is he or she also likely to show high intraindividual variability in other cognitive domains? Contrary to expectation, there was not a clear pattern of positive interrelationships among the intraindividual variability estimates for the three cognitive measures. The findings suggest that, unlike levels of performance, intraindividual variability may not always be a consistent phenomenon cutting across different domains of cognitive functioning. Instead, the intraindividual variability observed for a specific cognitive ability seems to be independent of intraindividual variability in other cognitive domains. It is important to note that this lack of relationship is not simply because of unreliability or noise variance in the intraindividual variability indexes. Within a particular measure, the intraindividual variability within one block of occasions was generally positively related with intraindividual variability within other blocks of occasions, suggesting a reasonably high level of interindividual stability of intraindividual variability (i.e., trait-like consistency). One way to think about this is that each measure might have its own reliable intraindividual variability signature.

The dissimilarity between this finding and those of previous studies may stem from the measures that were used in the current investigation. Evidence of positive intraindividual variability correlations across different measures has typically come from investigations that used latency scores derived from reaction time tasks (Anstey, 1999; Hultsch et al., 2002; MacDonald et al., 2003; Rabbitt et al., 2001; Strauss et al., 2002). Therefore, the high intercorrelations among intraindividual variability estimates among these measures may reflect task similarity or perhaps a common underlying cognitive process, such as psychomotor speed. Our findings suggest that perhaps intraindividual variability in distinct, albeit related, domains of cognitive functioning as assessed by psychometric measures is not strongly associated. Similar results were reported by Li et al. (2001), who found that performance variability on measures assessing different aspects of memory (e.g., short-term memory, text recall, spatial memory) was not strongly interrelated.

Another possible explanation for low between-measure correlations in intraindividual variability may stem from the fact that, in this study, intraindividual variability was related to practice-related improvements in performance. The cognitive tests studied (reasoning, memory, speed) did not evince the same pattern of practice-related improvement, differing in both the magnitude and linearity of their practice-related trajectories. Thus, differential practice effects may have yielded differential intraindividual variability effects. Regardless of the underlying cause, this finding has potentially important implications for the study of intraindividual variability in older adults' psychological functioning. Specifically, it appears that one cannot assume that intraindividual variability for measures of different cognitive domains will covary or that it is meaningful to seek some global cognitive intraindividual variability estimate that is predictive of other outcomes. Hence, future studies of intraindividual variability will most likely not be able to develop a single parsimonious model that explains intraindividual variability across all domains of cognition, but instead

will have to focus on a more multidimensional description and explanation of intraindividual variability.

As noted earlier, there was substantial trait-like consistency of intraindividual variability within each measure; across blocks, correlations were generally positive and modest to strong, particularly for list memory and perceptual speed. However, for the inductive reasoning task, initial analyses suggested that something closer to a quasi-simplex pattern (Gorsuch, 1974) was observed, in which intraindividual variability was strongly related between adjacent blocks, but virtually no correlation between the earlier and later blocks was observed. However, follow-up analyses revealed the presence of participants performing at ceiling levels in the latter occasions of the study. Once ceiling participants were removed from the analyses, the correlations between Block 1 and the last two blocks were similar to those of the other two cognitive measures. These results are consistent with previous work (Hultsch et al., 2002; Nesselroade & Salthouse, 2004; Rabbit et al., 2001) suggesting that, within a particular measure, intraindividual variability tends to be a consistent and reliable property and not a reflection of random error or noise.

The third question of this study addressed the extent to which mean performance and intraindividual variability were interrelated. As mentioned, prior theoretical and empirical research has generally viewed intraindividual variability as a vulnerability index, with increased levels reflecting difficulty in maintaining homeostatic self-regulation. Therefore, persons who have the lowest performance levels (most vulnerable in terms of low performance) would also have greater intraindividual variability (most vulnerable in terms of inconsistency), a finding that has been well supported in the cognitive aging literature (Anstey, 1999; Hultsch et al., 2000, 2002, MacDonald et al., 2003; Rabbit et al., 2001). Results from this investigation indicated that higher mean performance on the cognitive measures was associated with greater intraindividual variability in performance on those measures.

In an attempt to explore the connection between practice-related improvement and intraindividual variability, exploratory analysis further examined the correlation between individual slope parameters and intraindividual variability within each block. Although not completely uniform, the results suggested that individual differences in practice-related improvement within a block of occasions were positively associated with intraindividual variability. That is, even after controlling for within-person practice-related improvement, the magnitude of observed variability was, in general, positively associated with individual differences in the slope of practice-related improvement within each block of 30 occasions. This finding was particularly strong for the inductive reasoning task, which exhibited the largest linear growth.

Taken together, our findings suggest that when practice-related gain is present, higher mean performance is associated with greater intraindividual variability, which is consistent with the results from studies examining intraindividual variability in the earlier portion of the life span (Siegler, 1994; van Geert & van Dijk, 2002) and more recent characterizations of the intraindividual variability as a multidimensional construct (Li et al., 2004). In fact, in the child development literature, intraindividual variability of performance is often considered an indicator of nascent improvement, reflecting the active exploration and use of differing performance strategies when performing new complicated tasks or, as Siegler (1994) has stated, the overlapping, or the ebbing and flowing, of new and old ways of thinking. A similar finding from the aging literature is offered by Li et al. (2001), who reported significant and positive relationships between intraindividual mean and variability for two spatial memory tasks and a negative relationship for two measures of text recall.

Siegler (1994; Siegler & Lemaire, 1997) and Li et al. (2001, 2004) offered a dual interpretation of intraindividual variability in older adults' cognitive performance (e.g., adaptive vs. maladaptive intraindividual variability) to account for the differing mean-variability relationship. Whether adaptive or maladaptive intraindividual variability is observed depends, in part, on the cognitive measures used in a particular study. Maladaptive intraindividual variability is typically observed for reaction time-based tasks generally thought to reflect psychomotor speed, which are unlikely to evince substantial growth or improvement. Consequently, for tasks in which individuals are not expected to improve much with practice, intraindividual variability assumes its expected interpretation as inconsistency or failure of homeostasis. Similar findings are also found in studies examining intraindividual variability in noncognitive domains that do not show improvement with repeated testing such as walking (Li et al., 2001), control (Eizenman, Nesselrode, Featherman, & Rowe, 1997), self-efficacy (Lang, Featherman, & Nesselrode, 1997), and health and activity (Ghisletta, Nesselrode, Featherman, & Rowe, 2002). Adaptive intraindividual variability will most likely be observed for cognitive tasks in which individuals show substantial improvement as a function of repeated assessment and may reflect individuals' active testing of new performance strategies. This type of intraindividual variability is expected to be observed in accuracy-based measures, like those in the current study, in which individuals have more degrees of freedom with respect to performance and, therefore, a greater potential for variability. However, as noted by Li et al. (2004), once acquisition or improvement in performance has ended and individuals reach asymptotic levels of performance, intraindividual variability represents maladaptive processes.

A limitation of the current work is that we did not include an assessment of strategy usage. Nonetheless, it does seem reasonable to speculate that some of the variance in the intraindividual variability exhibited by participants reflected their active engagement in generating and discarding strategies. In fact, the cognitive abilities included in this investigation have been used in intervention studies (Ball et al., 2002), which specifically focus on training participants in the strategies needed to improve performance (Saczynski, Willis, & Schaie, 2002). Li et al. (2001) offered a similar interpretation for their spatial recall task, stating that it was more prone to self-generated strategy use than the other memory measures. Future research will benefit from including assessments of participants' strategy use to determine if in fact varying degrees of successful strategy use are related to intraindividual variability in performance.

Although our results add to the extant knowledge of intraindividual variability in cognitive performance, they must be considered alongside a number of limitations. The present sample was quite homogeneous. The majority of the participants were relatively well-educated Caucasian elders who reported a high mean income. Consistent with other longitudinal studies, the results from the attrition analysis indicated that the participants who completed the study were better educated and had higher levels of education and income than those who dropped out before completion of the study. These differences, coupled with the fact that completers were older than persons who dropped out, suggest that our sample was particularly selective and may have been composed of relatively "successful agers." In addition, there are most likely unassessed personality and motivational differences between completers and dropouts, which may also be related to the results of the current investigation. Second, although the self-administrative nature of the daily assessment was a unique design aspect separating the current investigation from previous intraindividual variability studies, it also introduced a lack of experimental oversight over the daily testing situation. Although the gain, from our perspective, is an unparalleled density of observations, the cost is clearly to experimental control. For instance, within- and between-person differences in adherence to the time limitations of the cognitive tests may have served to influence the magnitude of intraindividual variability observed. Even a deviance of

a few seconds to a time limit can add an extra item or two. Although a number of strategies were built into the testing protocol to minimize cheating (e.g., signing of a declaration of noncollaboration, weekly data reviews), the self-administration protocol used means that we cannot affirm quantitatively that our participants were perfectly faithful to the study protocol. However, we did observe strong relationships and no mean differences between the observed and the mean of the first five unobserved measurement occasions. Ultimately data quality was an issue of trust, although all participants did sign the declaration of noncollaboration on all the workbooks. It is important to note that the potential costs to data fidelity, in our view, were offset by the unprecedented frequency and duration of the testing occasions, which could have only been accomplished through self-administration. Although self-administration has been widely used in affect and social diary studies (Almeida, McGonagle, Cate, Kessler, & Wethington, 2003; Charles & Pasupathi, 2003; Eaton & Funder, 2001) with great success, the implementation of home-based computerized administration seems like a reasonable middle-ground strategy to limit concerns about protocol fidelity.

In addition, unlike latency-based measures of reaction time, the paper-and-pencil performance-based measures used in the current study did not protect against the presence of ceiling effects. Clearly, future research examining intraindividual variability in performance on psychometric ability measures will need to use computerized adaptive testing, or at least measures with larger ranges of difficulty, to ensure the fidelity of the testing procedure and to minimize ceiling and floor performance.

Although the concept of intraindividual variability in psychological domains is not new (Fiske & Rice, 1955), the gerontological literature has seen a dramatic growth in the theoretical and empirical interest in variability in older adults. This investigation represents an addition to this area of aging research and establishes a meaningful connection with studies examining intraindividual variability in the earlier portion of the life span. Given its modern infancy, there is still much to be discovered about the meaning and patterns of variability in older adults' cognitive performance. However, a multidimensional and multidirectional approach that embraces both maladaptive and adaptive intraindividual variability in old age will be central to the further development of this area of inquiry.

Acknowledgments

Partial financial support for this study was provided by the Association for Gerontology in Higher Education, AARP Andrus Foundation Doctoral Fellowship in Gerontology, and the Wayne State University's Vice President for Research. This article is based on the doctoral research of Jason C. Allaire, which was conducted under the direction of Michael Marsiske.

References

- Almeida DM, McGonagle KA, Cate RC, Kessler RC, Wethington E. Psychosocial moderators of emotional reactivity to marital arguments: Results from a daily diary study. *Marriage and Family Review* 2003;34:89–113.
- Anstey KJ. Sensorimotor variables and forced expiratory volume as correlates of speed, accuracy and variability in reaction time performance in late adulthood. *Aging, Neuropsychology, and Cognition* 1999;6:84–95.
- Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, et al. Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association* 2002;288:2271–2281. [PubMed: 12425704]
- Baltes, PB.; Reese, HW.; Nesselroade, JR. *Life-span developmental psychology: Introduction to research methods*. Mahwah, NJ: Erlbaum; 1977.

- Bruhn P, Parsons OA. Reaction time variability in epileptic and brain-damaged patients. *Cortex* 1977;13:373–384. [PubMed: 414884]
- Burton CL, Hultsch DF, Strauss E, Hunter MA. Intraindividual variability in physical and emotional functioning: Comparison of adults with traumatic brain injuries and healthy adults. *Clinical Neuropsychologist* 2002;16:264–279. [PubMed: 12607140]
- Charles ST, Pasupathi M. Age-related patterns of variability in self-descriptions: Implications for everyday affective experience. *Psychology and Aging* 2003;18:524–536. [PubMed: 14518813]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Erlbaum; 1988.
- de Weerth C, van Geert P. Changing patterns of infant behavior and mother–infant interaction: Intra- and interindividual variability. *Infant Behavior and Development* 2002;24:347–371.
- de Weerth C, van Geert P, Hoijtink H. Intraindividual variability in infant behavior. *Developmental Psychology* 1999;35:1102–1112. [PubMed: 10442878]
- Eaton LG, Funder DC. Emotional experience in daily life: Valence, variability, and rate of change. *Emotion* 2001;1:413–421. [PubMed: 12901400]
- Eizenman DR, Nesselroade JR, Featherman DL, Rowe JW. Intraindividual variability in perceived control in an older sample: The MacArthur successful aging studies. *Psychology and Aging* 1997;12:489–502. [PubMed: 9308096]
- Ekstrom, RB.; French, JW.; Harman, H.; Derman, D. *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service; 1976. (rev. ed.)
- Fiske DW, Rice L. Intra-individual response variability. *Psychological Bulletin* 1955;52:217–250. [PubMed: 14371891]
- Ford, DH. *Self-constructing living systems: A developmental perspective on behavior and personality*. Hillsdale, NJ: Erlbaum; 1987.
- Fuentes K, Hunter MA, Strauss E, Hultsch DF. Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome. *Clinical Neuropsychologist* 2001;15:210–227. [PubMed: 11528543]
- Ghisletta P, Nesselroade JR, Featherman DL, Rowe JW. Structure and predictive power of intraindividual variability in health and activity measures. *Swiss Journal of Psychology* 2002;61:73–83.
- Gorsuch, RL. *Factor analysis*. Philadelphia: W. B. Saunders; 1974.
- Hendrickson, AE. The biological basis of intelligence: Part I. Theory. In: Eysenck, HJ., editor. *A model for intelligence*. Berlin: Springer-Verlag; 1982. p. 151-196.
- Hultsch DF, MacDonald SWS, Dixon RA. Variability in reaction time performance of younger and older adults. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 2002;57:101–115.
- 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]
- Lang FR, Featherman DL, Nesselroade JR. Social self-efficacy and short-term variability in social relationships: The MacArthur Successful Aging Studies. *Psychology and Aging* 1997;12:657–666. [PubMed: 9416633]
- 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]
- Li SC, Huxhold O, Schmiedek F. Aging and attenuated processing robustness: Evidence from cognitive and sensorimotor functioning. *Gerontology* 2004;50:28–34. [PubMed: 14654724]
- Li, SC.; Lindenberger, U. Cross-level unification: A computational exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In: Nilsson, L-G.; Markowitsch, HJ., editors. *Cognitive neuroscience of memory*. Kirkland, WA: Hogrefe & Huber; 1999. p. 103-146.
- 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]

- Martin M, Hofer SM. Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology* 2004;50:17–21. [PubMed: 14654722]
- Nesselroade, JR. The warp and woof of the developmental fabric. In: Downs, RM.; Liben, LS.; Palmero, DS., editors. *Visions of aesthetics, the environment and development: The legacy of Joachim F. Wolhwill*. Hillsdale, NJ: Erlbaum; 1991. p. 213-240.
- Nesselroade, JR.; Ford, DH. Methodological considerations in modeling living systems. In: Ford, ME., editor. *Humans as self constructing living systems: Putting the framework to work*. Hillsdale, NJ: Erlbaum; 1987. p. 47-79.
- Nesselroade JR, Jones CJ. Multi-modal selection effects in the study of adult development: A perspective on multivariate, replicated, single-subject, repeated measures designs. *Experimental Aging Research* 1991;17:21–27. [PubMed: 1936098]
- 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:P49–P55.
- Orsega Smith E, Payne LL, Godbey G. Physical and psychosocial characteristics of older adults who participate in a community-based exercise program. *Journal of Aging and Physical Activity* 2003;11:516–531.
- Piek JP. The role of variability in early motor development. *Infant Behavior and Development* 2002;25:452–465.
- 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: Human Experimental Psychology* 2001;54:981–1003.
- Rey A. L'examen psychologique dans les cas d'encephalopathie traumatique [Psychological testing in cases of traumatic brain injury]. *Archives de Psychologie* 1941;28:21.
- Rittle Johnson B, Siegler RS. Learning to spell: Variability, choice, and change in children's strategy use. *Child Development* 1999;70:332–348. [PubMed: 10218258]
- Saczynski JS, Willis SL, Schaie KW. Strategy use in reasoning training with older adults. *Aging, Neuropsychology, and Cognition* 2002;9:48–60.
- Salthouse TA. Attentional blocks are not responsible for age-related slowing. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 1993;48:P263–P270.
- Schaie, KW.; Hofer, SM. Longitudinal studies in aging research. In: Birren, JE.; Schaie, KW., editors. *Handbook of the psychology of aging*. 5. San Diego: Academic Press; 2001. p. 53-77.
- Schmidt, M. *Rey Auditory Verbal Learning Test: A handbook*. Los Angeles: Western Psychological Services; 1996.
- Siegler RS. Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science* 1994;3:1–5.
- Siegler RS, Lemaire P. Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General* 1997;126:71–92. [PubMed: 9090145]
- Singer T, Verhaeghen P, Ghisletta P, Lindenberger U, Baltes PB. The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging* 2003;18:318–331. [PubMed: 12825779]
- Strauss E, MacDonald SWS, Hunter M, Moll A, Hultsch DF. Intraindividual variability in cognitive performance in three groups of older adults: Cross-domain links to physical status and self-perceived affect and beliefs. *Journal of the International Neuropsychological Society* 2002;8:893–906. [PubMed: 12405540]
- Stuss DT, Pogue J, Buckle L, Bondar J. Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. *Neuropsychology* 1994;8:316–324.
- Thurstone, TG. *Primary mental ability for grades*. Chicago: Science Research Associates; 1962. p. 9-12.(rev. ed.)
- van Dijk M, de Goede D, Ruhland R, van Geert P. Kindertaal met bokkensprongen. Een studie naar intra-individuele variabiliteit in de taalontwikkeling [Child language cuts capers: A study into

intraindividual variability in developmental processes]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden* 2001;56:26–39.

van Geert P, van Dijk M. Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development* 2002;25:340–374.

West R, Murphy KJ, Armilio ML, Craik FIM, Stuss DT. Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition* 2002;49:402–419. [PubMed: 12139961]

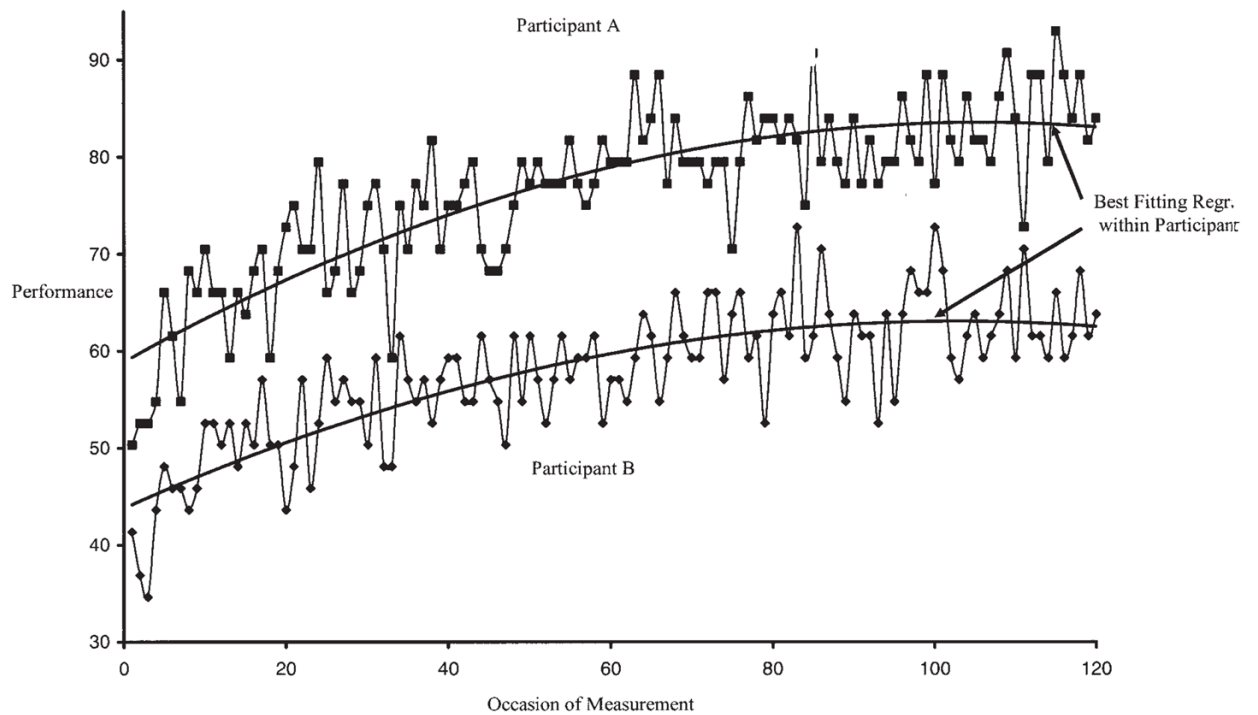


Figure 1. Best fitting regression line (Regr.) within a high-performing participant (A) and a low-performing participant (B).

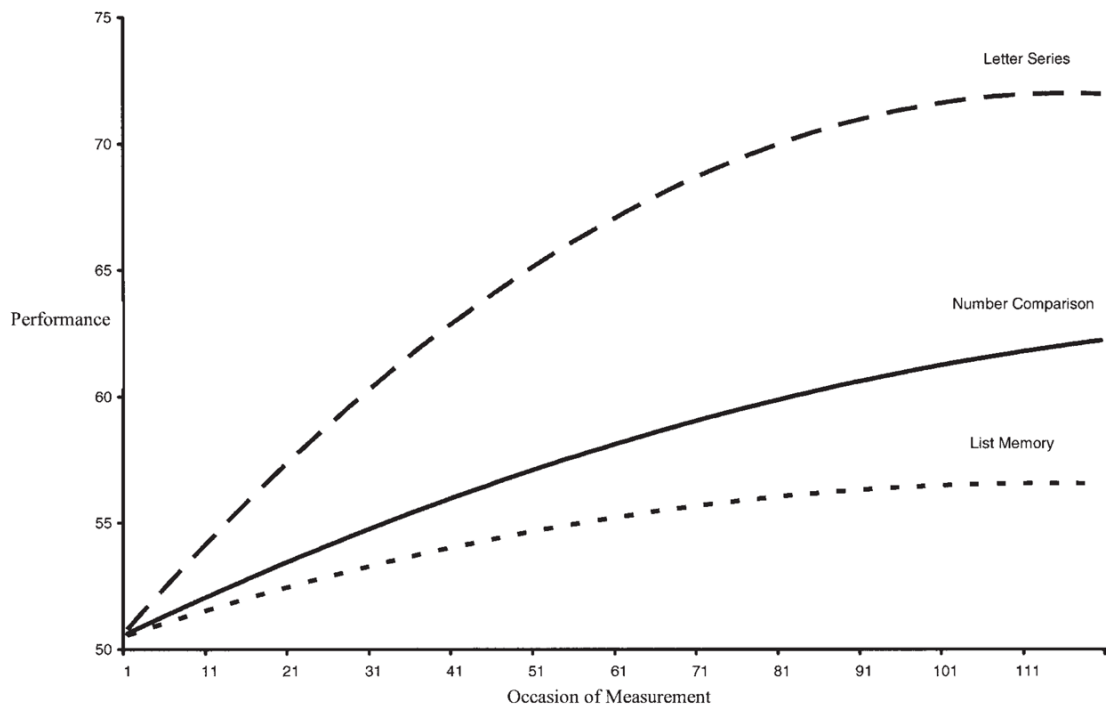


Figure 2. Average growth trends for the Letter Series Test, list memory task, and Number Comparison Test over the 120 occasions of repeated measurement.

Table 1

Description of Cognitive Measures Included in the Daily Mental Exercise Workbooks

Cognitive test	Hypothesized ability	Adapted from	No. of items (time allowed)	Sample item
Letter Series Test	Inductive reasoning	Thurstone (1962)	30 (4 min)	Complete the series: <i>a m b a n b a _</i>
Number Comparison Test	Perceptual speed	Ekstrom, French, Harman & Derman (1976)	100 (90 s)	Compare pairs of number strings ranging from 3 to 13 digits; mark an <i>x</i> between pairs that are the same: <i>1 5 7 8 _ 1 5 6 8</i> <i>3 2 7 _ 3 2 7</i>
List memory	Memory	AVLT (Rey, 1941)	12 (1-min study, 1-min recall)	Recall a list of 12 common words in any order e.g., (<i>mother house</i>)

Table 2
 Mean Differences Between Participants Who Dropped Out (n = 14) and Who Completed the Study (n = 36)

Variable	Dropout		Completers		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Age	70.50	5.85	73.92	5.51	-1.94	.06
Education	14.07	2.37	16.36	2.98	-2.57	.01
Income	11.53	3.45	14.43	2.33	-3.33	.00
Letter Series Test	44.79	10.90	52.03	8.99	-2.41	.02
Number Comparison Test	46.26	9.48	51.45	9.49	-1.68	.10
List memory	48.48	11.41	50.59	9.51	-.67	.51

Table 3

Covariation Among Intraindividual Variability Over Blocks (n = 36)

Variable	Block	Block			
		1	2	3	4
Letter Series Test	1. Block 1	—	—	—	—
	2. Block 2	.46*	—	—	—
	3. Block 3	.09	.32	—	—
	4. Block 4	.18	.47*	.54*	—
Number Comparison Test	1. Block 1	—	—	—	—
	2. Block 2	.48*	—	—	—
	3. Block 3	.38*	.35*	—	—
	4. Block 4	.59*	.56*	.58*	—
List memory	1. Block 1	—	—	—	—
	2. Block 2	.74*	—	—	—
	3. Block 3	.82*	.77*	—	—
	4. Block 4	.68*	.79*	.66*	—

* $p < .05$.

Table 4

Correlation Between Variability and Mean Performance Within Block (n = 36)

Block	Letter Series Test	Number Comparison Test	List memory
1	.69* (.76*)	.26	.40* (.42*)
2	.24 (.52*)	.30	.44* (.44*)
3	-.22 (.15)	.20	.47* (.43*)
4	-.10 (.36)	.33*	.23 (.42*)

Note. Correlations in parentheses represent relationships after removing ceiling participants (Letter Series $n = 29$; list memory $n = 31$).

* $p \leq .05$.

Table 5Correlation Between Variability and Slope Within Blocks ($n = 36$)

Block	Letter Series Test	Number Comparison Test	List memory
1	.54* (.59*)	-.06	.24 (.22)
2	.42* (.48*)	.45*	.29 (.18)
3	.25 (.07)	-.09	-.10 (.13)
4	.32* (.26)	.14	.22 (.29)

Note. Correlations in parentheses represent relationships after removing ceiling participants (Letter Series Test $n = 29$; list memory $n = 31$).

* $p \leq .05$.

Appendix

Descriptive Statistics for Cognitive Measures

Cognitive test	120 Occasions		Block 1		Block 2		Block 3		Block 4	
	M	SD	M	SD	M	SD	M	SD	M	SD
Letter Series Test	68.48	13.03	59.89	11.05	68.73	13.21	73.39	14.20	76.07	0.07
Number Comparison Test	54.59	8.69	51.96	8.01	54.27	8.20	55.85	9.51	56.56	10.59
List memory	57.59	9.28	52.65	7.83	56.64	9.65	59.16	10.14	61.75	10.89