Neuropsychology
2005, Vol. 19, No. 1, 77–87

0894-4105/05/$12.00 DOI: 10.1037/0894-4105.19.1.77

Correlates of Within-Person (Across-Occasion) Variability in Reaction Time

Timothy A. Salthouse and Diane E. Berish
University of Virginia

A total of 420 adults between 18 and 91 years of age carried palm pilot (Palm M100; PalmOne, Milpitas, CA) devices and performed several reaction time trials when prompted at random times during the day. On the average, within-person variability in median reaction time from one occasion to the next was nearly the same magnitude as the between-persons variability in across-occasion mean reaction time. Analyses controlling the variation in one variable when examining relations involving the other variable suggested that an individual's mean reaction time is a more powerful predictor of relations with age and with the level of various cognitive abilities than his or her across-occasion *SD*.

In recent years there has been growing interest in within-person variability as a potentially informative individual difference parameter. At least five factors may contribute to this interest. First, high levels of within-person variability may signify health-related problems. In fact, Hultsch, MacDonald, Hunter, Levy-Bencheton, and Strauss (2000) recently suggested that "intraindividual variability may be a behavioral indicator of compromised neurological mechanisms" (p. 588), and greater within-person variability of cognitive functioning compared with healthy individuals has been documented in a number of patient groups (e.g., Hetherington, Stuss, & Finlayson, 1996; Hultsch et al., 2000; Murtha, Cismaru, Waechter, & Chertkow, 2002; Stuss, Pogue, Buckle, & Bondar, 1994).

Second, unusual levels of within-person variability might function as an early indicator of impending cognitive change. That is, large fluctuations in one's momentary level of cognitive performance may be a precursor to certain types of cognitive pathology, and information about variability might provide a more useful baseline against which to evaluate the severity of extreme behaviors. The occurrence of periods when performance is very low could also be important because in some circumstances, a critical aspect of functioning is the minimal level of performance rather than the maximum level that can be achieved on a particular occasion. Moreover, there is some evidence that the strongest relations with other cognitive abilities are with measures of an individual's worst performance rather than his or her best, or average, performance (Coyle, 2003).

Third, large within-person variability (i.e., low across-occasion consistency) could distort the evaluation of an individual's level of cognitive functioning. For example, in clinical assessments, the existence of substantial within-person variability could result in the same individual being considered normal on some occasions and impaired on other occasions. More generally, the phenomenon of within-person variability raises the question of the degree to which cognitive performance should be considered to reflect stable trait-like abilities as opposed to being influenced by transient states.

Fourth, within-person variability could contribute to inconsistency in research results across studies because in a single-occasion study, it is impossible to distinguish relatively stable trait variance from fluctuating state variance. For example, if there are age differences in within-person variability, it could lead to spurious conclusions of age-related increases in between-persons variability (Nesselroade, 2001). The possibility of large within-person variability in the performance of neuropsychological tests would also raise questions about the basis of the correspondence typically assumed between test performance and brain function.

Finally, identification of correlates of within-person variability may be informative about possible causes of individual differences, and particularly age-related individual differences, in cognitive functioning. That is, moment-to-moment variation in cognitive performance within an individual may be analogous to naturally occurring moderators of age-related effects, in which case they might be as informative as other moderators in discovering causes of those effects.

Two types of within-person variability can be distinguished. Within-occasion variability refers to the variation from one measurement (or trial) to another measurement when the measurements are separated by seconds to minutes within the same occasion. In contrast, across-occasion variability refers to the variation in performance across different occasions or sessions that are separated by intervals ranging from minutes to weeks. The primary interest in the current article is in across-occasion variability because the aggregation across multiple measurements within the same occasion is likely to result in assessments that are less influenced by unsystematic factors (i.e., measurement error), and as the intervals between assessments increase, it is more likely that each occasion corresponds to a distinct fluctuation in state.

At least three recent studies have investigated within-person variability in samples of older adults. In each study, several cognitive tasks were administered to older adults at weekly intervals for between 4 and 36 weeks (i.e., Hultsch et al., 2000; Li, Aggen,

Timothy A. Salthouse and Diane E. Berish, Department of Psychology, University of Virginia.

This research was supported by National Institute of Aging Grant RO1AG19627 to Timothy A. Salthouse. We thank Karen Siedlecki for helping coordinate the data collection and John Nesselroade and Nilam Ram for valuable comments on drafts of the article.

Correspondence concerning this article should be addressed to Timothy A. Salthouse, Department of Psychology, University of Virginia, Charlottesville, VA 22904. E-mail: salthouse@virginia.edu

Nesselroade, & Baltes, 2001; Rabbitt, Osman, Moore, & Stollery, 2001). All three studies were similar in finding considerable within-person (across-occasion) variability and in reporting that participants who exhibited large variability in one of the variables tended to exhibit large variability in other variables. Another common finding across the studies was the existence of significant correlations between the within-person variability in the repeated variable and the level of performance in other cognitive variables, in all cases in the direction of higher cognitive functioning among adults with lower levels of within-person variability.

Within-person variability was investigated in adults across a wider age range by Nesselroade and Salthouse (2004). In their project, adults between 18 and 91 years of age performed two different tasks on three separate sessions. One of the tasks was a visual–motor tracking task in which the participant manually controlled a track ball to try to keep a cursor centered on a moving target. The other task involved participants drawing lines to connect targets as rapidly as possible in either a continuous sequence (i.e., either numerical or alphabetical) or in an alternating (i.e., number-letter-number, etc.) sequence.

The primary analyses consisted of computing a mean for each person across the four or five trials in each session and then computing the standard deviation (*SD*) of each person's means across the three sessions. A similar pattern was apparent with each variable, and it can be described by focusing on the variable representing the speed of connecting targets in alternating numeric and alphabetic sequence. The between-persons *SD* for this variable was 1.95, and the average within-person (across-session) *SD* was 1.04. The ratio of these two values is .53, which indicates that the average across-session variation for a given person was about 53% as large as the between-persons variation apparent across adults ranging from 18 to 91 years of age. Because the slope (in *SD* units per year) of the regression equation relating mean performance to age was .04, it can be inferred that the average within-person variability for this variable was roughly equivalent to the variation expected across a period of 26 years of normal aging (i.e., $1.04/.04 = 26$).

The results of the Nesselroade and Salthouse (2004) study, together with those from the studies of Hultsch et al. (2000), Li et al. (2001), and Rabbitt et al. (2001), clearly indicate that within-person variability is large relative to between-persons variability and that it is significantly related to age and to a variety of cognitive variables. Furthermore, the fact that the studies involved different cognitive tasks, different samples, and different numbers of and intervals between occasions suggests that the phenomenon of large within-person variability in cognitive functioning is quite robust.

The phenomenon of large within-person variability is interesting in its own right, but because measures of variability are often highly correlated with measures of central tendency, it is also important to attempt to determine the degree to which parameters corresponding to central tendency and to variability have independent relations with age and with other cognitive variables. There are at least two possibilities for how variability and central tendency might be related to one another. One possibility is that occasional lapses of attention increase the variability, which in turn might affect the central tendency for time or error measures. A second possibility is that if the variable has a lower boundary, such as is the case with time or error measures, then an upward shift in the distribution of scores will be associated with a greater distance from the lower boundary, thereby providing more opportunity for increased variability. In this case, it would be increases in the central tendency that are responsible for increases in variability. These two interpretations differ with respect to which parameter is considered primary in their relation to one another, but they both view the two parameters as closely related to one another, and thus it is desirable to attempt to distinguish their unique contributions to relations with other variables.

Two methods have been used to attempt to separate the influence of the central tendency (i.e., mean) from analyses involving variability (i.e., *SD*). The indirect method involves partialling correlates of the mean from scores on individual trials before computing the measure of variability. Hultsch et al. (2000) used this technique when they partialled age and practice effects from individual reaction times (RTs) before computing variability and then examined the relation of age to this residual variability estimate. This method is indirect because it does not actually control the variance in the mean when examining relations involving the *SD* but instead only controls the variance in variables that are correlated with the scores on which the mean is based. The method results in the elimination of the relation of the partialled variables to the mean, but if the relation between the mean and *SD* is not altered, then this method does not resolve the question of which parameter is critical for the relations with other variables.

The more direct method of dealing with the relation between central tendency and variability consists of statistical control of either the central tendency parameter or the variability parameter when examining relations involving the other parameter. Because the parameters are highly correlated, a hierarchical regression method of statistical control can be used to minimize problems of collinearity (cf. Cohen & Cohen, 1983, p. 115). An advantage of the direct analytical method is that it is possible to examine unique relations of each parameter by controlling the *SD* when examining relations involving the mean and then controlling the mean when examining relations involving the *SD*. A finding of an asymmetric reduction in the relations with other variables would suggest that the variable associated with the greatest reduction may be more fundamental in terms of the relations with other variables. Salthouse (1993b, 1998) used this analytical procedure in two separate projects with measures of within-occasion within-person variability in reaction time and found that the relations of age and cognitive variables with the mean were only slightly reduced when the *SD* was controlled but that there was a large reduction in the relations with the *SD* when the mean was controlled. These results therefore suggest that, at least for within-occasion variability, the central tendency may be more fundamental than variability in terms of the relations with the level of other variables.

Because relevant data were available in the Nesselroade and Salthouse (2004) project, analyses similar to those just described were conducted on these data by first controlling the *SD* when examining relations involving the mean and then controlling the mean when examining relations involving the *SD*. Several points should be noted about the results of these analyses, which are summarized in Table 1. The first is that the means and within-person *SD*s for each variable were positively correlated, with the correlations ranging from .56 to .90. The second point is that both the means and the *SD*s were significantly related to age (positively) and to a variety of cognitive variables (negatively). However, the

Table 1
Results of Reanalyses of Data From Nesselroade and Salthouse (2004)

Predictor	*M*	*M.SD*	*SD*	*SD.M*
Tracking error[a]				
Age	.34*	.26*	.22*	−.04
Reas/Space1	−.50*	−.33*	−.37*	−.03
Reas/Space2	−.50*	−.30*	−.40*	−.02
Memory	−.39*	−.33*	−.22*	.09
Speed	−.36*	−.28*	−.24*	.03
Vocabulary	−.28*	−.20*	−.20*	.00
Tracking lag[b]				
Age	.27*	.17	.22*	.09
Reas/Space1	−.37*	−.36*	−.13	.09
Reas/Space2	−.40*	−.43*	−.09	.20*
Memory	−.31*	−.23*	−.22*	−.05
Speed	−.36*	−.28*	−.23*	−.04
Vocabulary	−.21*	−.21*	−.08	.05
Connections—same[c]				
Age	.59*	.60*	.37*	−.38*
Reas/Space1	−.47*	−.51*	−.27*	.34*
Reas/Space2	−.53*	−.56*	−.31*	.37*
Memory	−.44*	−.52*	−.24*	.37*
Speed	−.67*	−.73*	−.39*	.49*
Vocabulary	−.05	−.17	.03	.17
Connections—alternating[d]				
Age	.33*	.30*	.21*	−.14
Reas/Space1	−.64*	−.39*	−.51*	.05
Reas/Space2	−.62*	−.36*	−.51*	.03
Memory	−.47*	−.30*	−.37*	.05
Speed	−.51*	−.37*	−.36*	.12
Vocabulary	−.36*	−.12	−.35*	−.09

Note. Entries in the *M* column are correlations with the mean, those in the *M.SD* column are semipartial correlations with the mean after controlling the *SD,* those in the *SD* column are correlations with the *SD,* and those in the *SD.M* column are semipartial correlations with the *SD* after controlling the mean. Reas/Space1 = average of *z* scores for block design and analysis synthesis; Reas/Space2 = average of *z* scores for Raven's, spatial relations, and paper folding; Memory = average of *z* scores for logical memory, free recall, and paired associates; Speed = average of *z* scores for digit symbol and cross-out; Vocabulary = average of *z* scores for the Wechsler Adult Intelligence Scale Vocabulary Test and Woodcock-Johnson Picture Vocabulary Test.
[a] Correlation between mean and *SD* is .72. [b] Correlation between mean and *SD* is .56. [c] Correlation between mean and *SD* is .90. [d] Correlation between mean and *SD* is .84.
* $p < .01$.

most important point is that with each variable, the relations involving the mean were only slightly reduced when the *SD* was controlled, but the relations involving the *SD* were either eliminated or reversed when the mean was controlled.[1]

Although the results in Table 1 are consistent with those reported by Salthouse (1993b, 1998), they are nevertheless surprising because they suggest that within-person variability may be less fundamental than a measure of the individual's average performance in terms of relations with other variables. However, the amount of within-person variability may have been underestimated in this and other prior studies because participants scheduled the testing occasions themselves, and many of the occasions likely occurred at the same time each day. It is therefore possible that within-person variability could be greater and might exhibit stronger relations with other variables when the occasions occur at unpredictable times during the day. This possibility was investigated in the current project by asking participants to carry a palm pilot (Palm M100; PalmOne, Milipitas, CA) during their normal daily activities for a period of 3 to 7 days. The device was programmed to sound an alarm at random times, and on each of these occasions (or probes) the participant performed a number of trials of a choice RT task. The participants also performed a variety of cognitive tests in the laboratory, and thus it was possible to examine the relations of composite measures of cognitive ability to each individual's mean RT and his or her *SD* (computed from the scores across probe occasions).

Study 1

Method

Participants. Palm pilot devices (Palm M100) were assigned to 164 adults who participated in a larger study in which they reported to the laboratory for three sessions (Salthouse, Atkinson, & Berish, 2003). Several of the devices failed to operate correctly ($n = 5$), some participants apparently did not fully understand the task because their mean RTs were greater than 3 s ($n = 2$), and some participants responded on fewer than three occasions ($n = 8$); thus, analyzable data were only available from 149 participants. Characteristics of these 149 adults, arbitrarily divided into three age groups, are summarized in Table 2. Inspection of the table reveals that there was little relation of age to self-rated health, years of education, or age-adjusted scaled scores for the vocabulary and digit symbol variables but that increased age was associated with more of a morning preference, a higher level of vocabulary ability, and lower levels of several cognitive abilities. All of the participants were volunteers primarily recruited through advertisements and referrals from other participants. Because the age-adjusted scaled scores for the vocabulary and digit symbol variables were higher than the means of the nationally representative normative samples, the participants can be considered to be high functioning.

Procedure. The palm pilot task was controlled with the Experience Sampling Program (*ESP: The Experience Sampling Program*, http://www2.bc.edu/~barretli/esp/). Participants were instructed on how to operate the palm pilot device and then completed one probe in the presence of the examiner. They were encouraged to respond to as many probes as possible except while sleeping, driving, or engaged in other activities where responding would be disruptive. However, because the probes occurred randomly while participants were engaged in their normal activities, there was no record of what they were doing at the time of the probe.

The palm pilots were programmed to present 105 probes across a 7-day interval, which corresponded to about one probe every 96 min, because one probe occurred in every epoch, which was defined as the total number of minutes for the interval (i.e., $7 \times 24 \times 60$) divided by the number of probes (i.e., 105). Because there were no time restrictions on when the probes were presented, about one third of them were probably presented when the participants were sleeping. Furthermore, 50% of the participants completed the three sessions of the primary study in less than 7 days; hence, the palm pilot was returned before all probes had been presented. These factors make it difficult to determine exactly how many probes could have been responded to by each participant, but the number of probes with data ranged from 3 to 92, with a mean across participants of 21.7. Nearly all of the participants who were asked about their experience with the palm pilot reported that it was not disruptive of their routines and was rather enjoyable.

[1] An additional set of analyses was conducted after partialling age from the scores at each occasion. As one would expect, this resulted in the elimination of the relation between age and the across-occasion mean, but there were still substantial correlations between the mean and the *SD* (i.e., a range from .52 to .85).

Table 2
Participant Characteristics (N = 149) in Study 1

Variable	Age range (years) 18–39	40–59	60–91	All	*r*
N	43	68	38	149	
Age	28.3 (6.4)	49.5 (4.9)	69.6 (5.7)	48.5 (16.3)	
% Women	46.5	71.0	68.4	63.1	.19*
Self-rated health	2.0 (0.7)	2.1 (0.9)	1.8 (0.8)	2.0 (0.8)	−.01
Years of education	15.7 (2.7)	15.9 (2.3)	16.2 (2.5)	15.9 (2.5)	.12
Morningness–eveningness	44.9 (10.8)	58.6 (9.6)	60.1 (9.5)	55.2 (11.7)	.51*
Vocabulary scaled score	12.6 (3.2)	12.8 (2.6)	12.7 (2.2)	12.7 (2.7)	.03
Digit symbol scaled score	11.9 (3.2)	11.9 (3.1)	12.3 (2.3)	12.0 (2.9)	.04
Reas/Space	0.48 (0.82)	0.09 (0.70)	−0.63 (0.72)	.02 (0.85)	−.50*
Memory	0.27 (0.86)	0.14 (0.71)	−0.50 (0.61)	.01 (0.79)	−.37*
Speed	0.60 (0.80)	0.03 (0.70)	−0.70 (0.63)	.01 (0.86)	−.59*
Vocabulary	−0.47 (1.02)	0.21 (0.73)	−0.15 (0.78)	.00 (0.88)	.29*
Executive functioning	0.32 (0.49)	0.11 (0.52)	−0.45 (0.66)	.03 (0.62)	−.46*
No. of probes with responses	17.7 (8.5)	22.1 (12.5)	27.7 (16.2)	21.7 (12.9)	.24*
Mean reaction time (RT)	0.84 (.18)	1.05 (.25)	1.21 (.27)	1.03 (.27)	.52*
Across-occasion *SD*	0.18 (.13)	0.27 (.27)	0.31 (.17)	0.26 (.22)	.19*
RT reliability	0.61 (.20)	0.54 (.21)	0.48 (.21)	0.54 (.21)	−.23*

Note. Variables in parentheses are between-persons standard deviations. Health was rated on a scale ranging from 1 (*excellent*) to 5 (*poor*). Morningness–eveningness preference was assessed with a questionnaire (Horne & Ostberg, 1976). The vocabulary and digit symbol scaled scores are age-adjusted scaled scores, which in the normative data (from the WAIS III) have means of 10 and standard deviations of 3. The reas/space composite was the average of *z* scores for the Raven's progressive matrices, paper folding, and spatial relations variables. The memory composite was the average of *z* scores for the logical memory, free recall, and paired associates variables. The speed composite was the average of *z* scores for the WAIS III digit symbol, letter comparison, and pattern comparison variables. The vocabulary composite was the average of *z* scores for the WAIS III Vocabulary Test, Woodcock-Johnson Picture Vocabulary Test, and synonym and antonym variables. The executive functioning composite was the average of *z* scores for the Wisconsin Card Sorting Test number of errors (reverse coded), Tower of Hanoi number of moves (reverse coded), time difference between connecting items in alternating versus same sequence (reverse coded), and number of items generated in verbal fluency and figural fluency variables. WAIS = Wechsler Adult Intelligence Scale.
* $p < .01$.

The reaction time task involved the display of a number between 1 and 5 in the middle of the screen, and five boxes numbered 1–5 on the bottom of the screen (see the top panel of Figure 1 for an illustration of the display in this task). Participants were instructed to use the palm pilot stylus to tap the box corresponding to the displayed stimulus number as quickly as possible. On every probe occasion the five digits were each presented twice in random order, and RT was recorded in 1/100 of a second. To minimize the influence of extreme scores, the median RT across the 10 trials served as the measure of RT at each occasion.

The participants also reported to the laboratory on 3 separate days during which they performed a battery of cognitive tests. The variables used in the current analyses are briefly described in Table 3. Please refer to the study by Salthouse, Atkinson, and Berish (2003) for more complete information about the variables and participants. On the basis of the factor analysis described in that report, composite cognitive variables were formed by averaging *z* scores for the variables loading on each factor.

Results[2]

For each individual, the median of the 10 RTs at each occasion was determined, and the mean and *SD* of these median RTs across occasions was computed.[3] The between-persons average of the mean RTs was 1.03 s, and the between-persons *SD* was .27 s. These values are quite long for a choice RT task, and this may be attributable to a combination of perceptual factors, as the stimuli were rather small and were displayed in relatively low contrast and of response factors, as the RTs included the time to lift, move, and tap the stylus.

Nevertheless, the RTs appear to be valid as measures of speed, because the mean RT was moderately correlated with score on the Wechsler Adult Intelligence Scale III (WAIS III; Wechsler, 1997a) Digit Symbol Test (i.e., $r = -.58$), which is often considered a measure of perceptual speed, but were not significantly correlated (i.e., $r = -.10$) with the score on another type of test, the WAIS III (Wechsler, 1997a) Vocabulary Test, which is assumed to measure word knowledge.

A coefficient alpha estimate of the reliability of the RTs at each occasion was obtained for each individual by treating the 10 RTs at each occasion as items and considering the occasions as replicates (analogous to individuals). In the current context, this statistic essentially evaluates the degree to which the RTs at a given occasion are more similar to one another than they are to the RTs on other occasions. The average coefficient alpha was .54, indicating that for most participants, the RTs within an occasion were

[2] Because of the large number of statistical comparisons and the moderately large sample sizes, a .01 level of statistical significance was used in all analyses.

[3] Preliminary inspection of the data revealed that errors were very rare; thus, they were ignored in all subsequent analyses. Furthermore, although for many participants there was a slight decrease in median RT across occasions, the magnitude of the across-occasion change in median RT was relatively small and was not related to across-occasion variability. It was therefore ignored in subsequent analyses.

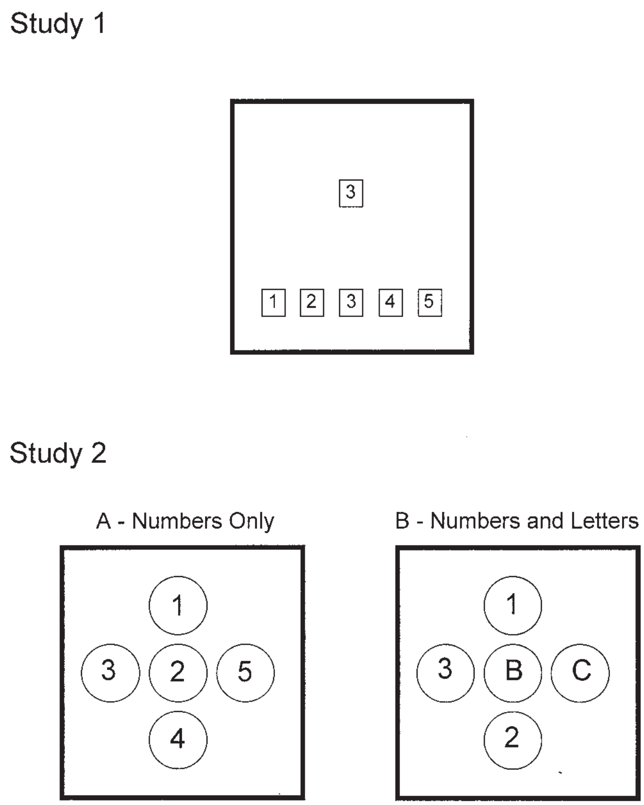

Figure 1. Illustration of the palm pilot displays for the reaction time tasks in Study 1 (top) and Study 2 (bottom). In Study 1, the task was to use a stylus to tap the box in the bottom row containing the digit that matched the digit in the middle of the screen. In Study 2, the numbers only task required the participant to tap the surrounding circle that contained the number following the number in the center circle. The numbers and letters task was similar except that the items were to be tapped in alternating numeric and alphabetic sequence (i.e., 1-A-2-B-3-C, etc.).

more closely related than were the RTs across occasions. This estimate of reliability was not significantly correlated with the RT mean ($r = -.18$) or with the RT *SD* ($r = .08$).

An illustration of the within-person variability in RT is portrayed in Figure 2, which displays the median RT at each occasion for a male participant who was near the middle of both the age distribution (i.e., 52 years old) and the distribution of mean (across-occasion) RTs (i.e., $M = 1.01$ s). The estimated reliability of his RTs was also near the middle of the values for the participants in this study (i.e., .62). However, it should be noted that this individual was actually somewhat less variable across occasions than the average participant, because his across-occasion *SD* was 0.19 compared with the mean for the sample of 0.26. The horizontal axis in Figure 2 represents time, with the vertical lines corresponding to the transition from 1 day to the next (i.e., 12 a.m.). Filled circles represent probes with responses, and unfilled circles indicate probes for which there was no response. Inspection of the data in the figure reveals that there was considerable fluctuation in the median RTs across occasions, with a range from about 0.7 s to 1.6 s.

The labels on the right vertical axis of Figure 2 indicate the average age corresponding to a given RT based on the parameters of a linear regression equation relating (across-occasion) mean RTs to age. Reference to this axis reveals that on some occasions, this person's median RT was close to the average of individuals who were 20 years old, whereas on other occasions, his or her median RT was close to the average of adults who were 70 years old.

In an attempt to determine whether some of the across-occasion variability in RT was related to when the probe occurred, correlations were computed for each participant between the median RT for a given occasion and the time in the day (on a 0–2,400 scale) at which that occasion occurred. The mean correlation across participants between median RT and the time of the probes was .01, indicating that there was no overall relationship between median RT and the time of the probe. Most of the responses occurred between about 8 a.m. and 8 p.m., and the average correlation between median RT and probe time was also very low for responses occurring within this interval. Because people might have differed in their relationship between probe time and median RT, correlations were also computed between an individual's time–RT correlation and his or her age and self-reported morningness–eveningness preference (Horne & Ostberg, 1976). Neither of these correlations was significant, as the age correlation was −.14 and the morningness–eveningness correlation was .00.

The results described thus far indicate that the average within-person (across occasion) variability was very large relative to the between-person variability in mean RT. Perhaps because the occurrence of the occasions was unpredictable, the within-person variability relative to the between-person variability was even greater in this study than in the Nesselroade and Salthouse (2004) study. In fact, the mean within-person *SD* of .26 s was nearly the same magnitude as the between-persons *SD* in the mean RTs (.27 s).

The second major question in this study was whether it was the mean or the *SD* that was primarily responsible for relations with age and other cognitive variables. This question was investigated by examining zero-order correlations and semipartial correlations in which either the *SD* or the mean is controlled by means of hierarchical regression when considering relations involving the other variable.[4] Because both the means and *SD*s were related to age, the analyses were repeated after also controlling the variation in age.

Results of these analyses are presented in the top panel of Table 4. The initial row indicates that the correlation between the RT mean and the RT *SD* was .76. The remaining rows indicate that the mean RT was moderately correlated with age and with all of the composite cognitive variables except for the vocabulary composite. The correlations were smaller with the *SD* variable, but several were significantly different from zero. Most importantly, when the *SD* was controlled there was only a slight reduction in correlations involving the mean, but when the mean was controlled the correlations with the *SD* were either eliminated or were reversed in

[4] The statistical control analyses remove only the linear relations between the variables; thus, the nature of the relation between the mean and *SD* was examined with polynomial terms in hierarchical regression. The R^2 associated with the linear relation between RT and *SD* was .57, that for the quadratic relation was .05, and that for the cubic relation was .00. Corresponding values for the RT A variable in Study 2 were .44, .01, and .00, and those for the RT B variable in Study 2 were .62, .01, and .00, respectively. Because the linear trend was responsible for the largest proportion of the relation between the variables, linear regression techniques were considered appropriate for purposes of statistical control.

Table 3
Description of Cognitive Variables Included in Analyses

Variable	Description	Source
Reas/space		
Spatial relations	Determine the correspondence between a 3-D figure and alternative 2-D figures	Bennett et al. (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)
Raven's	Determine which pattern best completes the missing cell in a matrix	Raven (1962)
Episodic memory		
Logical memory	Number of idea units recalled across three stories	Wechsler (1997b)
Free recall	Number of words recalled across Trials 1–4 of a word list	Wechsler (1997b)
Paired associates	Number of response terms recalled when presented with a stimulus term	Salthouse et al. (1996)
Speed		
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)
Vocabulary		
WAIS	Provide definitions of words	Wechsler (1997a)
Picture	Name the pictured object	Woodcock & Johnson (1990)
Antonym	Select the best antonym of the target word	Salthouse (1993a)
Synonym	Select the best synonym of the target word	Salthouse (1993a)
Executive functioning		
WCST	Number of errors in the Wisconsin Card Sorting Test (WCST)	Heaton et al. (1993)
Tower	Number of moves taken to solve 3-disk Tower of Hanoi problem	Neth (2002)
Verbal fluency	Generate as many words as possible within 60 s that begin with a specified letter	
Figural fluency	Generate as many unique patterns by drawing lines between dots	Ruff (1996)
Connect difference	Difference between connecting lines with nonalternating sequences (i.e., numbers or letters) and connecting lines with alternating sequences (i.e., numbers–letters, letters–numbers)	Salthouse et al. (2000)

Note. Reas/space = reasoning and spatial visualization ability; 3-D = three dimensional; 2-D = two dimensional; WAIS = Wechsler Adult Intelligence Scale.

direction. All of the correlations were reduced when the variance in age was controlled; however, the same pattern was still evident of a slight reduction in the correlations involving the mean after control of the *SD*, but either elimination or reversal of the correlations involving the *SD* after control of the mean. Another set of analyses also controlled the reliability estimate when examining relations between the mean or *SD* and the age and composite cognitive variables, and the pattern of correlations was nearly identical to that in Table 4.

Discussion

The results of the analyses summarized above provide clear answers to the two questions motivating this study. First, the average within-person (across-occasion) variability was very large and in fact was nearly the same magnitude as the variability apparent across people in their mean (across-occasion) RT. Because, depending on the occasion that happens to be sampled, a given individual may perform at a level near the average of the best or the worst performer in the sample, this finding has important implications for assessment and for inferences about brain functioning.

The results with respect to the second question concerning relations of the mean and *SD* to other variables were also unambiguous. The data in this study suggest that the relations with other variables appear to be primarily attributable to relations with the mean (at least when it represents the central tendency of observations across multiple occasions) and not to across-occasion variability around the mean. This finding is consistent with results of the reanalyses of the data reported in Nesselroade and Salthouse (2004), and with the earlier results on within-occasion variability reported in Salthouse (1993b, 1998).

The failure to find strong relations between median RT and time of the probe appears to contradict the idea that there are strong circadian variations in performance. However, recent research on time-of-day effects in cognitive performance has been somewhat inconsistent because the effects appear to vary according to the age

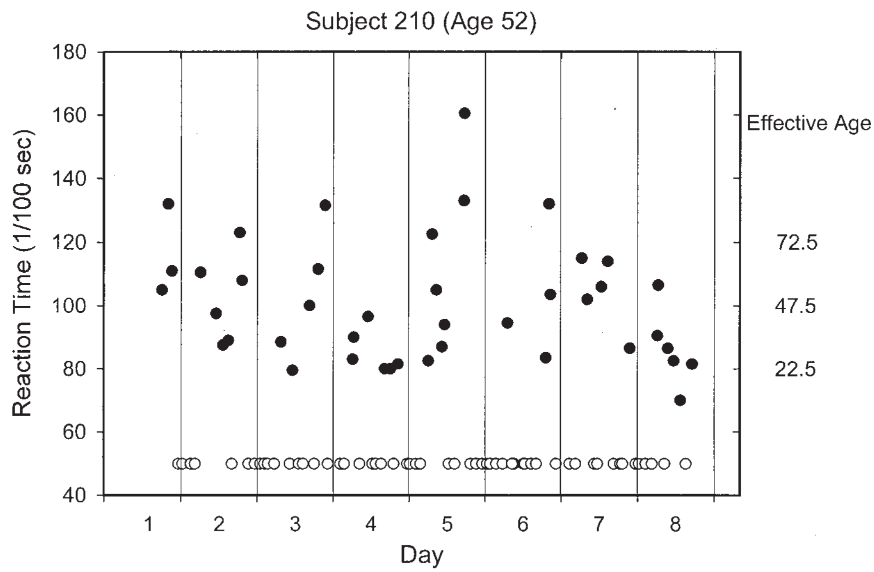

Figure 2. Median reaction time (RT) at each occasion for one participant in Study 1. Each filled circle represents a probe with responses, and each unfilled circle represents a probe occasion at which the individual did not respond. The vertical axis on the right of the graph indicates the average age corresponding to the RT on the right axis based on parameters of a linear regression equation relating mean RT to age across the 149 participants in the study.

of the participant and the type of task, and no main effects of time-of-day were reported in several studies (e.g., Brown, Goddard, Lahar, & Mosley, 1999; Li et al., 1998; Salthouse et al., 2003; West, Murphy, Armilio, Craik, & Stuss, 2002).

Study 2

The primary goal of Study 2 was to attempt to replicate and extend the results of Study 1. One modification of the procedure was based on the possibility that some of the within-person variability may be due to the extreme range of probe times, which could occur at any time during the day or night. Although on the average there was no systematic relation between probe time and median RT, some participants had very slow RTs early in the morning or late at night, and these abnormally long RTs may have inflated the estimates of across-occasion variability and possibly distorted relations with other variables. Probes in the current study were therefore presented only between 8 a.m. and 8 p.m.

Another modification from Study 1 was the use of different RT tasks based on Parts A and B of the Trail Making Test to determine whether the pattern of large within-person variability and stronger relations with the mean than with the *SD* also hold with more complex RT tasks. In an attempt to determine whether some of the within-person variability in RT might be associated with current mood, the participants also rated their mood at each occasion. Finally, to determine whether an explicit payment would increase compliance in responding to the probes, participants received $1 for each probe with a response.

Method

Participants. Participants were recruited in the same manner and from the same pool as in Study 1 but were tested in 2003 instead of 2002. More details about the participants and the other tasks performed during the three laboratory sessions are reported in the study by Salthouse, Berish, and Siedlecki (in press). Palm pilots were assigned to a total of 291 participants, but 11 malfunctioned, 3 participants apparently did not fully understand the task because their mean RTs in the simpler (RT A) condition were greater than 3 s, and 6 participants responded to fewer than three probes. Characteristics of the remaining sample of 271 adults are summarized in Table 5. It can be seen that in most respects the sample is similar to that in Study 1.

Procedure. The palm pilots were programmed to present six probes between 8 a.m. and 8 p.m. for each of 5 consecutive days. Probes therefore occurred once every 120 min for 12 hr per day for up to 5 days (depending on when the participant returned for his or her third laboratory session).

Immediately prior to the performance of the RT tasks, on each occasion the participants used the palm pilot to rate their current mood. The mood ratings were based on a 6-point scale ranging from 1 (*very sad*) to 6 (*very happy*). Although measurement based on a rating of a single item is not ideal, the average mood rating was significantly correlated (−.41) with depression as measured by the Center for Epidemiological Studies—Depression (Radloff, 1977) and (−.37) with the personality trait of neuroticism (Costa & McCrae, 2003).

On each probe occasion there were two types of RT trials with displays similar to those illustrated in the bottom of Figure 1. Notice that each display consisted of a center circle with four surrounding circles located above, below, to the right, and to the left of the center. In the A Trials (numbers only), only numbers appeared in the circles, and the task was to tap the circle containing the number that followed the number in the center circle. To illustrate, in the example in the lower left of Figure 1, the left circle should be tapped next because it contains the number following the number currently in the center circle. When a response was registered, a new display immediately appeared, with the next number in the sequence from the previous display now in the center and a different arrangement of numbers in the surrounding circles. This sequence continued for five displays with starting numbers varying randomly between 1 and 5.

Table 4

Correlations With the Mean and the Standard Deviation of RTs Before and After Control (via Hierarchical Regression) of Age and the Other Variable

Variable	*M*	*M.SD*	*M*.Age	*M*.Age.*SD*	*SD*	*SD.M*	*SD*.Age	*SD*.Age.*M*
Study 1								
Age	.52*	.57*			.19*	−.30*		
Reas/space	−.39*	−.38*	−.16	−.14	−.19*	.17	−.09	.05
Memory	−.34*	−.40*	−.17	−.24*	−.10	.24*	−.02	.17
Speed	−.61*	−.61*	−.36*	−.37*	−.28*	.28*	−.17	.18*
Vocabulary	−.02	.06	−.20	−.14	−.08	−.10	−.14	.03
Executive functioning	−.44*	−.42*	−.24*	−.21*	−.22*	.17	−.14	.08
Study 2, A Trials (numbers only)								
Age	.59*	.45*			.39*	−.01		
Reas/space	−.53*	−.35*	−.30*	−.20*	−.40*	−.07	−.24*	−.07
Memory	−.45*	−.37*	−.30*	−.27*	−.26*	.05	−.14	.05
Speed	−.57*	−.45*	−.30*	−.26*	−.35*	.04	−.14*	.04
Vocabulary	−.02	.01	−.27*	−.20*	−.04	−.04	−.20*	−.04
Study 2, B Trials (numbers and letters)								
Age	.42*	.41*			.22*	−.18*		
Reas/space	−.44*	−.38*	−.26*	−.22*	−.26*	.14	−.16*	.07
Memory	−.36*	−.33*	−.23*	−.22*	−.20*	.14	−.13	.10
Speed	−.47*	−.38*	−.25*	−.19*	−.30*	.12	−.18*	.04
Vocabulary	−.04	.03	−.21*	−.13	−.08	−.07	−.16*	.01

Note. The column labeled *M* contains correlations with the mean; the column labeled *M.SD* contains semipartial correlations with the mean after controlling the *SD*; the column labeled *M*.Age contains semipartial correlations with the mean after controlling age; the column labeled *M*.Age.*SD* contains semipartial correlations with the mean after controlling for age and *SD*. The same labeling convention applies to the *SD* columns on the right of the table. Higher reaction times (RTs) correspond to slower performance, and thus a negative correlation between the RT mean or *SD* and a cognitive variable indicates that people with faster or less variable RTs had higher cognitive scores. Reas/space = reasoning and spatial visualization ability.
* $p < .01$.

The B Trials (numbers and letters) were similar except that the circles contained either letters or numbers, and the task was to tap circles in alternating numeric and alphabetic sequence. To illustrate, in the example in the lower right of Figure 1, the left circle should be tapped because it contains the number that follows the letter in the center circle. Five trials of a given type were presented in succession, but whether the five A Trials occurred before or after the five B Trials varied randomly across occasions. The median of the five RTs of each type served as the performance measure on each occasion.

Results

The overall mean of the mood ratings across participants and occasions was 4.5, which corresponds to a rating of moderately happy. The between-persons *SD* of the mean (across-occasion) mood ratings was 0.6, and the mean within-person (across-occasion) *SD* was 0.8. These values indicate that the variation in mood from one occasion to the next for a given individual was somewhat larger than the between-persons variation in average mood rating.

As expected, the RTs were much slower with the tasks in this study than with the task in Study 1 (i.e., $M = 1.03$ s), and the RTs in the B Trials were much slower than those in the A Trials (i.e., 3.67 s vs. 1.29 s). Although there were only 5 repetitions of each task at each occasion instead of 10 as in the prior study, the average estimated reliabilities in the A (i.e., .55) and B (i.e., .56) Trials were nearly the same as that in Study 1 (i.e., .54). The estimated reliabilities for the A and B Trials had correlations of .13 and .09 with the means and correlations of .24 and .36 with the *SD*s.

In both RT tasks, the within-person variability was nearly the same magnitude as the between-persons variability in average performance. That is, for the A Task, the average within-person across-occasion *SD* was .40, and the between-persons *SD* for the across-occasion means was .44. Corresponding values in the B Task were 1.85 and 1.77, respectively.

For each individual, correlations were computed across occasions between the median RTs for each type of trial, the rated mood, and the time of day at which the probe occurred. For most participants, the median RTs in the A and B Trials varied together because the mean across-occasion correlation was .35. However, the median RTs were not related to the time of day when the probe occurred (i.e., mean correlations of .01 for A and .00 for B) or to the rated mood (i.e., mean correlations of −.02 for A and .02 for B). Furthermore, although participants varied in the direction and magnitude of the relations between RT and time or rated mood, the within-person correlations were not related to age (i.e., correlations between .02 and −.15) or to morningness–eveningness preference (i.e., correlations from −.02 to .07). As in Study 1, therefore, the results of this study offer no evidence that RT variability was associated with time of day, and they add a new finding that there

Table 5
Participant Characteristics (N = 271) in Study 2

Variable	Age range (years) 18–39	40–59	60–89	All	*r*
N	101	119	51	271	
Age	28.0 (6.7)	49.6 (5.2)	70.0 (8.5)	45.4 (16.7)	
% Women	69.0	69.0	55.0	66.0	−.06
Self-rated health	1.9 (0.7)	2.1 (0.9)	2.0 (0.8)	2.0 (0.8)	.17*
Years of education	15.4 (2.4)	15.8 (2.8)	15.8 (2.3)	15.7 (2.5)	.08
Morningness–eveningness	47.2 (13.8)	54.4 (12.6)	57.6 (13.5)	52.3 (13.8)	.32*
Vocabulary scaled score	12.6 (3.2)	12.3 (2.8)	13.8 (2.1)	12.7 (2.9)	.05
Digit symbol scaled score	12.3 (2.9)	11.7 (2.8)	12.2 (2.8)	12.0 (2.9)	.01
Reas/space	0.39 (0.87)	−.02 (0.79)	−.65 (0.74)	.01 (.89)	−.48*
Memory	0.24 (0.74)	.03 (0.79)	−.49 (0.89)	.01 (.84)	−.35*
Speed	0.47 (0.68)	−.07 (0.69)	−.71 (0.72)	.01 (.81)	−.56*
Vocabulary	−0.41 (0.88)	.16 (0.89)	.45 (0.67)	.00 (.91)	.34*
No. of probes with responses	16.7 (7.1)	16.4 (7.2)	16.5 (6.9)	16.5 (7.1)	−.07
A Trials (numbers only)					
Mean reaction time (RT)	1.05 (.30)	1.31 (.36)	1.71 (.52)	1.29 (.44)	.59*
Across-occasion *SD*	.30 (.16)	.40 (.32)	.61 (.48)	.40 (.33)	.39*
RT reliability	.59 (.21)	.50 (.22)	.59 (.19)	.55 (.22)	−.06
B Trials (numbers and letters)					
Mean RT	3.05 (1.64)	3.66 (1.44)	4.93 (2.09)	3.67 (1.77)	.42*
Across-occasion *SD*	1.75 (1.39)	1.71 (0.94)	2.40 (1.29)	1.85 (1.21)	.22*
RT reliability	.59 (.21)	.52 (.21)	.58 (.20)	.56 (.21)	−.05

Note. Variables in parentheses are between-persons standard deviations. The reas/space composite was the average of *z* scores for the Raven's progressive matrices, paper folding, and spatial relations variables. The memory composite was the average of *z* scores for the logical memory, free recall, and paired associates variables. The speed composite was the average of *z* scores for the WAIS III Digit Symbol, letter comparison, and pattern comparison variables. The vocabulary composite was the average of *z* scores for the WAIS III Vocabulary Test, Woodcock-Johnson Picture Vocabulary Test, and synonym and antonym variables. WAIS = Wechsler Adult Intelligence Scale.
* $p < .01$.

was also no consistent relation with self-rated mood despite considerable across-occasion variability in both RT and rated mood.

The correlations involving the mean and the *SD* of the two RT variables are presented in the three panels of Table 4. It can be seen that the mean RTs had moderate correlations with age and with all composite cognitive variables except vocabulary and that these correlations were somewhat smaller but still significant when the *SD* was controlled. In contrast, the correlations with the *SD* were generally smaller than those with the mean, and statistical control of the mean eliminated all of the significant correlations involving the *SD* except for that with age, in which case the correlation was in the opposite direction, indicating that the *SD*s were smaller with increased age.

Discussion

There are two major results of these studies. The first is the demonstration of very large within-person variability in measures of RT. In two studies involving a total of 420 individuals, the median RT from one occasion to the next was found to vary as much as the mean (across occasion) RT varied among people who ranged from 18 to 91 years of age. Because the reliability estimates indicate that for most participants the within-occasion RTs were more similar to one another than were the between-occasions RTs, this across-occasion variability cannot merely be attributed to random fluctuation. Unfortunately, it is not yet clear what is responsible for the within-person variability, because it was not associated with the time during the day at which the probe occurred, even after taking the individual's time-of-day preference into consideration, or with the individual's self-rated mood on that occasion. It may be related to fluctuations in arousal or in the balance of hormones or other biochemicals, but these possibilities were not assessed in the present studies.

Regardless of its cause, the phenomenon of substantial across-occasion variability has important implications for the assessment of individuals because any single occasion can be considered to represent only one sample from a larger distribution of potential levels of performance that the individual could have exhibited. To the extent that other cognitive or neuropsychological variables also exhibit large within-person variability, sensitive evaluations in the future may need to incorporate multiple assessments across a period of days or weeks. Nesselroade (1991) described this type of multiple assessment as a measurement burst and particularly advocated its use in longitudinal designs. Although much more expensive and time-consuming than single-occasion assessments, measurement burst designs may be necessary to obtain the most sensitive and accurate indication of an individual's true level and range of cognitive performance.

Many neuropsychologists frequently assume that there is a direct correspondence between level of performance on a particular test and efficiency of functioning in a specific brain region, as implied by terms such as *frontal lobe tests*. However, the discovery that people vary substantially in their level of performance from one occasion to the next raises questions about the directness of, and basis for, this correspondence. Furthermore, to the extent that the fluctuations in performance are systematic, explanations are needed to account for the variations in brain functioning that are presumably responsible for the variations in performance.

The second major result of the present studies is the consistent finding that measures of within-person variability appear to be secondary to measures of central tendency with respect to relations with age and with a variety of cognitive variables.[5] For many variables, the mean and the *SD* are highly related because the variability around the mean is frequently greater when the mean is larger, but all of the analyses indicated that statistical control of the mean has a greater attenuating effect on correlations involving the *SD* than vice versa.

The differential pattern of relations with other variables does not necessarily imply that the mean is more fundamental than the *SD* in terms of their relations to one another. That is, the mean could be high because the variability is greater or variability could be large because the mean is high. However, the asymmetrical nature of the relations with other variables suggests that individual differences in the measure of central tendency are more important than individual differences in the measures of across-occasion variability with respect to relations with other individual difference characteristics such as age or the level of several cognitive abilities. Although this finding may appear surprising in light of the arguments outlined in the introduction and the magnitude of the observed within-person variability, the results appear to be robust because they were confirmed in two separate studies here as well as in the reanalyses of data from the Nesselroade and Salthouse (2004) project. Therefore, it appears that, at least on the basis of the results from this admittedly limited range of variables, if the measure of central tendency represents performance across multiple occasions, then it may carry more predictive power in samples of healthy adults than a measure of within-person variability in terms of relations with age and the level of various cognitive variables.

[5] Another possibility is that the estimates of variability were less precise, and perhaps less reliable, than were the estimates of central tendency. That is, if the variability parameter has more measurement error than the central tendency parameter, then it could result in the observed pattern of weaker relations with other variables because a smaller proportion of the variance is systematic and available to be associated with the other variables. Unfortunately, we are not aware of methods to compute precision or reliability for *SD*s, and thus this interpretation cannot be evaluated at the current time.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1997). *Differential Aptitude Test*. San Antonio, TX: Psychological Corporation.

Brown, L. N., Goddard, K. M., Lahar, C. J., & Mosley, J. L. (1999). Age-related deficits in cognitive functioning are not mediated by time of day. *Experimental Aging Research, 25,* 81–93.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Costa, P. T., & McCrae, R. R. (2003). *NEO-FFI (NEO Five-Factor Inventory)*. Lutz, FL: Psychological Assessment Resources.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31,* 567–587.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests.* Princeton, NJ: Educational Testing Service.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Odessa, FL: Psychological Assessment Resources.

Hetherington, C. R., Stuss, D. T., & Finlayson, M. A. J. (1996). Reaction time and variability 5 and 10 years after traumatic brain injury. *Brain Injury, 10,* 473–486.

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97–110.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14,* 588–598.

Li, K. Z. H., Hasher, L., Jonas, D., Rahhal, T. A., & May, C. P. (1998). Distractibility, circadian arousal, and aging: A boundary condition? *Psychology and Aging, 13,* 575–583.

Li, S-C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur Successful Aging Studies. *Gerontology, 47,* 100–116.

Murtha, S., Cismaru, R., Waechter, R., & Chertkow, H. (2002). Increased variability accompanies frontal lobe damage in dementia. *Journal of the International Neuropsychological Society, 8,* 360–372.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palermo (Eds.), *Views of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist, 6,* 187–193.

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual motor performance. *Journal of Gerontology: Psychological Sciences, 59B,* 49–55.

Neth, H. (2002). *MultiTowers*. Retrieved March 2002 from http://www.neth.de/Research/MultiTowers/MultiTowers_E. html

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *Quarterly Journal of Experimental Psychology, 54A,* 981–1003.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Raven, J. (1962). *Advanced progressive matrices, set II*. London: H. K. Lewis.

Ruff, R. M. (1996). *Ruff Figural Fluency Test*. Odessa, FL: Psychological Assessment Resources.

Salthouse, T. A. (1993a). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences, 48,* P29–P36.

Salthouse, T. A. (1993b). Attentional blocks are not responsible for age-related slowing. *Journal of Gerontology: Psychological Sciences, 48,* 263–270.

Salthouse, T. A. (1998). Relation of successive percentiles of reaction time distributions to cognitive variables and to age. *Intelligence, 26,* 153–166.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132,* 566–594.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776.

Salthouse, T. A., Berish, D. E., & Siedlecki, K. L. (in press). Construct validity and age sensitivity of prospective memory. *Memory & Cognition*.

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology, 10,* 272–285.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. (2000). Effects of aging on the efficiency of task switching in a variant of the Trail Making Test. *Neuropsychology, 14,* 102–111.

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. *Neuropsychology, 8,* 316–324.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Effects of time of day on age differences in working memory. *Journal of Gerontology: Psychological Sciences, 57B,* P3–P10.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Tests of Cognitive Ability, Revised*. Allen, TX: DLM Teaching Resources.

Received December 10, 2003
Revision received February 5, 2004
Accepted February 23, 2004 ■