Journal of Gerontology: PSYCHOLOGICAL SCIENCES
2004, Vol. 59B, No. 2, P49–P55

Methodological and Theoretical Implications of Intraindividual Variability in Perceptual-Motor Performance

John R. Nesselroade and Timothy A. Salthouse

Psychology Department, The University of Virginia.

As an individual differences variable, lability (within-person variability) has often been neglected even though it has been shown to predict key outcomes such as mortality. We examine intraindividual variability in perceptual-motor performance and relate it to chronological age in a sample of adults. The magnitude of between-session variability was found to average between 25% and 50% of the between-person variability and was equivalent in magnitude to the variation that was apparent across an age range of 12 to 27 years in cross-sectional comparisons. Age is related to the magnitude of intraindividual variability, which in turn is negatively related to performance on other cognitive tasks. Various implications of the findings are discussed.

SCIENTIFIC interest in short-term change and fluctuation in behavior has a long history (e.g., Cattell, 1957; Fiske & Rice, 1955; Thouless, 1936; Woodrow, 1932) and has increased substantially over the past two decades. In addition to numerous publications on substantive aspects of the topic (e.g., Butler, Hokanson, & Flynn, 1994; Eizenman, Nesselroade, Featherman, & Rowe, 1997; Hertzog, Dixon, & Hultsch, 1992; Hultsch & MacDonald, in press), treatments of pertinent methodological issues are also appearing with rapidity (e.g., Boker & Nesselroade, 2002; Browne & Nesselroade, in press; Hamaker, Dolan, & Molenaar, 2003; McArdle, 1982; McArdle & Hamagami, 2001; Molenaar, 1985; Moskowitz & Hershberger, 2002; Nesselroade & Molenaar, 1999; West & Hepworth, 1991). Evidence concerning the salience of intraindividual variability to the study of behavior and behavior development and change is becoming a compelling reminder that the prevailing emphasis on one of the seemingly most fundamental concepts in traditional differential psychology—stability of level of attributes across time—represents an oversimplification that can hinder the search for powerful and general lawful relationships (Nesselroade, 2002).

Just how important does information on intraindividual variability seem to be in the current state of behavioral inquiry? Behavioral scientists tend to answer that question in terms of magnitude of variation in one form or another (e.g., standard deviation units or effect size). Nearly 40 years ago, Bereiter (1963) reminded us that, in examining the intraindividual concept of "quotidian variability," Woodrow (1932) reported several instances in which between-occasions variability was considerably larger than between-persons variability on the same attributes. Bereiter (1963) suggested that the evidence of disparity between the magnitudes of the two kinds of differences—between persons and within persons—was an index of the relative inefficiency of individual-differences analysis as a substitute for studying intraindividual variability. When intraindividual variability in a given attribute is small, the interindividual differences in that attribute supply the bulk of the useful information, from a prediction standpoint; when intraindividual variability is large, however, they may not. Indeed, in the latter case, scores from only one occasion can yield highly misleading interindividual-differences information.

From the perspective of classical test theory, short term, intraindividual variability is a nuisance, albeit, in some cases, an "attractive" nuisance. Illustrative is Gulliksen's (1950) comment that the problem with estimating the reliability of a test from immediate test–retest correlations is that the estimates are too high because there "is no possibility for the variation due to normal daily variability to lower the correlation between forms" (p. 197). Opposing such negative sentiments are the more positive findings that short-term intraindividual variability is a valid indicator of substantively important events. Siegler (1994), for example, identified increased intraindividual variability in cognitive performance to be a "leading indicator" of impending cognitive change in children. Eizenman and associates (1997) described how individual differences in the magnitude of week-to-week intraindividual variability in perceived control were predictive of mortality a few years later. Indeed, as was pointed out by Rowe and Kahn (1987), elevated intraindividual variability on various attributes can be considered a risk factor for mortality in the elderly population.

More systematic investigation of the nature of intraindividual variability and change in a wide array of attributes is both compelling and timely. Here, we examine two distinct aspects of intraindividual, or within-person, variability in perceptual-motor performance. The first aspect on which we focus is methodological and is relevant to evaluating the representativeness of single-occasion assessment. The second aspect is theoretical and relates to whether there are age differences in moment-to-moment, or day-to-day, intraindividual variability and, if so, what are their salient features.

The Representativeness of a Single-Occasion Measurement

There are two principal ways to construe intraindividual variability. They are not mutually exclusive. The first is to regard it as essentially "noise" that, if necessary, can and should be dealt with in order to enhance any measured "signal." The second is to regard intraindividual variability as "signal" in its own right and to attempt to measure and explicate it. As was pointed out by a reviewer, one needs to distinguish between the nature of the intraindividual variability and the nature of individual differences in that intraindividual variability. The former could be "noise" in some basic sense (e.g., degraded neural capacity) even while the latter (individual differences in variability caused by differential degradation in neural capacity) could be predictive of other attributes and, therefore, definitely not "noise."

It is generally assumed that behavioral measurements reflect the following: (a) the construct of interest (e.g., ability or trait); (b) other irrelevant constructs; (c) short-term fluctuations that are due to shifts in arousal, motivation, and self-perception; and (d) otherwise unaccounted for errors of measurement. Because of the latter two kinds of influences especially, we can expect to find variability in repeated measurements taken on a given individual, and his or her scores on two different occasions may not be particularly close. If so, then which of the individual measurements is the "proper" one with which to characterize the individual? Obviously, neither may be. Trying to characterize the individual with a single score (except at a specific moment) may simply not be appropriate, however inconvenient this may be. Rather, it may be far more appropriate to use parameters of that intraindividual variability (e.g., the mean, the variance, amplitude, or latency) to characterize the individual and, by implication, differences among individuals.

If intraindividual variability is small relative to interindividual differences, it may be of only theoretical interest. However, if the former variation is large relative to the latter, and not simply error variance, it demands explanation and, as already noted, a mere examination of interindividual differences in level is not sufficient to apprehend the nature of the attribute in question.

As key as we believe it to be, however, intraindividual variability magnitude is difficult to interpret in absolute terms, so it is most usefully contrasted and compared with reference points such as other types of variability. Specifically, the magnitude of within-subject variability can be evaluated by expressing it relative to between-subject variability, or relative to the variation associated with an "interpretable" individual-difference variable such as age. At this historical point, intraindividual variability is unlikely to be important if the session-to-session variability is only a small fraction of the person-to-person variability or is equivalent to only a short span of normal aging. However, the implications for assessment are substantial if the estimates of within-person variation are large relative to between-person or across-age variation.

Age Differences in Intraindividual Variability

The second reason for an interest in intraindividual variability is theoretical, and it relates to whether there are age differences in moment-to-moment, or day-to-day, consistency. If so, what is the nature and significance of those age differences? For instance, is there a counterpart in older individuals of the findings reported by Siegler (1994) concerning increased intraindividual variability in performance indicating impending cognitive change in children? Is it possibly an early sign of decline when the level of cognitive performance fluctuates substantially in elderly persons?

Here is a more concrete illustration of the idea: Using the term "senior moments" to refer to temporary cognitive failures suggests an association between aging and cognitive fluctuation, but there has been little substantial research on this topic. An increase with age in intraindividual variability might be expected if fluctuating levels of performance are an early sign of cognitive decline. It is also possible that, for some variables, higher amounts of intraindividual variability in elderly persons are positive, rather than negative, outcomes. For example, higher variability might signify greater adaptability, less rigidity, or more creativity.

The literature reveals some attempts to rigorously evaluate the magnitude of intraindividual variability relative to inter-individual differences (e.g., Kraemer & Korner, 1976), but this is not a highly visible activity of past decades. Here, we assess the magnitude of both intraindividual variability and age-related differences in performance.

Methods

Participants

The total sample consisted of 204 participants ranging from 20 to 91 years of age. The participants were highly educated, averaging 16 years of formal education, and healthy, with an average rating of 2.1 (1 = excellent to 5 = poor). The average scores on two standardized cognitive measures (digit symbol and vocabulary) were substantially above the age-adjusted mean values (i.e., $M = 10$ and $SD = 3$) derived from a nationally representative normative sample. Participants were recruited through flyers, newspaper advertisements, and referrals from other participants. Cognitive tests were administered in three 2-hr laboratory sessions. More details on the sample and other tasks performed by the participants are contained in articles by Salthouse, Berish, and Miles (2002), Salthouse and Ferrer-Caja (2003), and Salthouse and Miles (2002). Key sample characteristics are presented in Table 1.

Data for this investigation were collected in two phases. Phase 1, which involved only the tracking task (described in the paragraphs that follow), included $n = 54$ adults, age 21 to 79. Phase 2, which involved both the tracking and connections tasks (also described in the paragraphs that follow), included $n = 150$ adults, age 20 to 91. For some analyses to be reported, participants from the two phases were pooled into one sample. For other analyses, they were kept separate.

Performance Tasks

Two perceptual-motor tasks were administered to the participants—a tracking task and a connections task. In the tracking task, a target (ball) moves randomly across the computer screen and the participant attempts to keep a cursor on the target by controlling a track ball with the preferred hand. Because the target and cursor positions are updated every 20 ms, continuous attention is required to perform well. The initial 20-s practice trial (not analyzed) is followed by five 70-s trials in each of the three sessions. Both tracking error and tracking lag are examined in the tracking task. The error is the root mean square error (RMSE) between target position and cursor position across the middle 60 s of the trial, and the lag is the

Table 1. Characteristics of the Pooled Sample

	Age Group		
Characteristic	18–39	40–59	60–91
n	52	84	68
Age (years)	27.9 (5.7)	49.5 (5.9)	70.0 (7.4)
Self-rated health	1.9 (0.8)	1.9 (0.9)	2.5 (0.8)
Years of education	15.5 (2.7)	16.2 (2.3)	16.1 (2.5)
Females (%)	75	70	53
Students (%)	25	0	0
WAIS BD scaled score	11.9 (3.9)	11.9 (3.2)	11.2 (3.3)
WAIS Voc scaled score	13.1 (3.6)	13.6 (2.8)	13.0 (3.5)

Notes: Standard deviations are given in parentheses. Health was rated on a scale ranging from 1 (excellent) to 5 (poor). The Vocabulary (Voc) and Block Design (BD) scaled scores are age-adjusted scores, which, in the normative data (from Wechsler Adult Intelligence Scale [WAIS] III), have $M = 10$ and $SD = 3$.

average delay in responding to the target (see Salthouse & Miles, 2002, for further description of these variables).

In the connections task (Salthouse et al., 2000), stimuli consisting of pages of 49 circles, with each circle containing either a number or a letter, are presented to the participant. The task is to draw lines to connect the circles as rapidly as possible according to numerical order, alphabetical order, alternating numerical and alphabetical, or alternating alphabetical and numerical order. Two pages of each condition were presented at each session, with 20 s allowed on each page. The score is the number of correct connections per page. For all subsequent analyses, the two same sequence conditions (i.e., numerical and alphabetical) were treated together; so were the two alternating sequence conditions (i.e., numerical and alphabetical; alphabetical and numerical).

Procedures

Participants were administered the same tasks on three separate sessions within a two-week period. Most of the sessions for a given individual were at approximately the same time of day. As already noted, the tracking task was performed for a practice trial and five additional trials at each session. There were four trials in the same condition and four trials in the alternating condition of the connections task on each session. Thus, the data permitted the examination of intraindividual variability both within and between sessions.

Results

Deriving Intraindividual Variability Scores for Individuals

Central to the analyses is the computation of scores derived from the basic data that represent individual participants' amounts of intraindividual variability. The construction of the average within-person variability and the between-person variability values are schematized in Figure 1 to further clarify their nature for the reader. As is generally the case with performance measures such as the ones used here, the means and standard deviations tend to be correlated. Over the four measures we report here, these correlations ranged from .56 to .90.

From the multisession, multitrial scores described herein, we

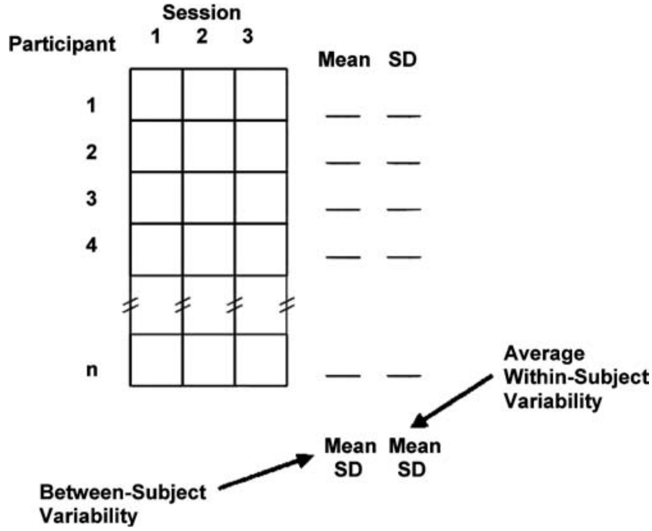

Figure 1. Schematic construction of within-person and between-person intraindividual variability.

computed two primary measures of within-person variability for each participant. First, the individual's standard deviation based on the within-session trial scores was computed for each of three sessions. These three standard deviations were then averaged. These values represent within-person, within-session (WPWS) variability. Second, the standard deviation of the individual's three mean session scores was computed. These values represent within-person, between-session (WPBS) variability.

These measures of within-person variability are standard deviations, but each is a score associated with an individual. Such scores were computed for four different measures: (a) tracking error; (b) tracking lag; (c) time per item to connect items in the same numeric or alphabetic sequence; and (d) time per item to connect items in alternating sequence. We elected to use standard deviations instead of variances because they are in the same metric as the original data and the means. We did not detrend the data or make other adjustments because the systematic changes in mean levels of performance across sessions were very small compared with the total between-session variability.

Characteristics of Intraindividual Variability

WPWS and WPBS variability scores were evaluated against two references. The first is the between-person variability (the standard deviation of the mean scores of individuals in the sample). The second reference for within-person variability scores is the slope of the cross-sectional age relation. If a variable is significantly related to age, the within-person standard deviation can be divided by the age slope to estimate the number of years of cross-sectional age difference corresponding to average within-person variability. The means and standard deviations across people (usual descriptive statistics) of the means and standard deviations for each session, and among the three sessions, are presented in Table 2. Also presented in Table 2 are the between-session correlations (test–retest stability coefficients) of the session means and standard

Table 2. Performance Means, Standard Deviations, and Between-Session Stability Coefficients

Session Information	Track Error	Track Lag	Same Sequence	Alternating Sequence
Session 1				
M[a]	70.8 (9.2)	305 (77)	0.90 (0.49)	2.40 (2.09)
SD[b]	5.1 (6.7)	23 (27)	0.27 (0.52)	1.67 (2.65)
Session 2				
M[a]	70.7 (10.8)	303 (72)	0.78 (0.51)	2.42 (2.32)
SD[b]	4.7 (5.3)	21 (27)	0.22 (0.78)	1.87 (2.96)
Session 3				
M[a]	69.7 (10.4)	300 (75)	0.72 (0.49)	2.24 (2.81)
SD[b]	4.2 (3.4)	22 (32)	0.19 (0.79)	1.42 (2.55)
Between sessions				
M	70.4 (9.2)	303 (70)	0.80 (0.40)	2.35 (1.95)
SD	3.8 (3.8)	26 (19)	0.17 (0.34)	1.04 (1.43)
Between-session stability coefficients: *M*				
Session 1–2 correlation	.72**	.82**	.58**	.51**
Session 1–3 correlation	.71**	.82**	.53**	.48**
Session 2–3 correlation	.76**	.81**	.26**	.46**
Between-session stability coefficients: *SD*				
Session 1–2 correlation	.20**	.70**	.39**	.35**
Session 1–3 correlation	.49**	.71**	.21*	.45**
Session 2–3 correlation	.58**	.71**	.02	.21**

Note: Standard deviations are given in parentheses.
[a]Average of the mean of all individual's scores on trials within the session.
[b]Standard deviation of individuals' trail means within the session.
*$p < .05$; **$p < .01$.

Table 3. Intraindividual Variability Characteristics and Relationships

Characteristic	Track Error	Track Lag	Connections: Same	Connections: Alternate
1. BP *SD*[a]	9.20	70.12	0.40	1.95
2. Average WPWS *SD*[b]	4.66	21.88	0.22	1.66
3. WPBS *SD*[c]	3.81	25.60	0.17	1.04
4. 2/3	1.22	0.85	1.29	1.60
5. Correlation of 2 with 3	.73**	.59**	.97**	.73**
6. 2/1	0.51	0.31	0.55	0.85
7. 3/1	0.41	0.37	0.43	0.53
8. r^2 of average WSIV with age	0.11	0.07	0.35	0.11
9. Slope of mean with age	0.18	1.08	0.014	0.038
10. 9/1	.020	.015	.035	.020
11. 2/9	25.89	20.26	15.71	43.68
12. 3/9	21.17	23.70	12.14	27.37

Notes: BP = between-person; WPWS = within-person, within-session; WPBS = within-person, between-session; WSIV = within-session intraindividaul variability. 4 = WSIV relative to between-session intraindividual variability (BSIV); 5 = correlation of WSIV with BSIV; 6 = WSIV relative to between-person variability; 7 = BSIV relative to between-person variability; 10 = annual cross-sectional age difference in between-person standard deviation units; 11 = WSIV relative to annual cross-sectional age differences; 12 = BSIV relative to annual cross-sectional age differences.
[a]Traditional measure of variation between participants.
[b]WPWS variation averaged over three sessions.
[c]Calculated as shown in Figure 1.
**$p < .01$.

deviations. These intersession correlations are noticeably higher for the tracking measures compared with the connections measures.

Table 3 summarizes several key characteristics of the intraindividual variability scores, including computations of within-person variability relative to between-person variability, and relative to estimated cross-sectional relations. The first row of Table 3 gives the between-persons standard deviations, one of our chief comparison references for evaluating intraindividual variability. These values represent what is ordinarily referred to as interindividual-differences information. The values in rows 2 and 3 are based on the two derived indices of intraindividual variability and show the magnitude of within-person variability—the average WPWS variability (row 2) and the WPBS variability (row 3). Rows 4 and 5 of Table 3 indicate that average WPWS intraindividual variability tends to be larger than WPBS intraindividual variability and that the two indices are substantially positively correlated—people who tend to vary more within sessions also tend to vary more across sessions.

Rows 6 and 7 of Table 3 compare the magnitudes of intraindividual variability to the magnitude of the between-persons variation. The ratios of rows 2 and 3 to 1 range between .31 and .85, with the majority being in the range from .40 to .55. By this method of calculation, the magnitude of intraindividual variability is about half of the between-persons variability. These values suffice to indicate that the amount of intraindividual variability is substantial.

All four measures of average within-session intraindividual variability are positively correlated with age, as shown in row 8 of Table 3. The WPWS intraindividual variability scores (means of the three within-session standard deviations) correlate consistently higher with age than the WPBS intraindividual variability scores. The reason may be as simple as the generally robust quality of means, but it may also reflect the tendency of the construction of the former scores to enhance stable individual differences. From a regression standpoint, these relationships represent modest to moderate performance variability–age slopes as shown in row 9, which in turn translate into relatively modest annual cross-sectional age differences in the metric of between-persons standard deviations as shown in row 10. The cross-sectional age trends are close to what are found with cognitive variables, which tend to range from –.02 to –.04 *SD*/year, or approximately 1 *SD* over 30–50 years (see, e.g., Salthouse & Ferrer-Caja, 2003; Salthouse et al., 2000).

The information in row 9 provides the basis for the second comparison reference described earlier—the slope of the cross-sectional age relation that can be used to estimate the number of years of cross-sectional age differences that correspond to average within-person variability. These values are presented in rows 11 and 12 of Table 3. Their implication is that the fluctuation for an individual from one occasion to the next is roughly equivalent to the variation apparent across people covering a span of 12 to 27 years.

We also computed a series of correlations between intraindividual variability indicators and Sex, Education, Gc (a composite measure of crystallized intelligence consisting of Wechsler Adult Intelligence Scale [WAIS] III vocabulary, see Wechsler, 1997a, and Woodcock–Johnson picture vocabulary, see Woodcock & Johnson, 1990, variables), Gf1 (a composite measure of fluid intelligence consisting of WAIS III block design, see Wechsler, 1997a, and Woodcock–Johnson analysis–

synthesis, see Woodcock & Johnson, 1990, variables), Gf2 (a second fluid ability composite made up of Raven's progressive matrices, see Raven, 1962, letter series, see Noll & Horn, 1998, spatial relations, see Bennett, Seashore, & Wesman, 1997, and paper folding, see Ekstrom, French, Harman, & Derman, 1976, variables), and Mem. Mem is a separate measure of specific memory based on story memory (Wechsler, 1997b), word list recall (Wechsler, 1997b), and paired associates recall (locally developed) variables (see Salthouse et al., 2002, for additional details). The results, summarized in Table 4, are generally quite consistent, showing intraindividual variability measures positively associated with age, unrelated to sex, and negatively correlated with education and nearly all of the cognitive variables (i.e., higher ability tends to be associated with less intraindividual variability).

Several conclusions can be reached in light of the values presented in Table 3. First, the session-to-session variability is considerable, averaging between .37 and .53 of the between-person standard deviation, and roughly equivalent to between 12 and 27 years of normal aging on these variables. Second, substantial intraindividual variability across a period of days implies poor precision of single-occasion assessments as estimates of fixed, true score values. Across-session variability could be noise, or it could be systematic. Third, although our evidence is not definitive, that people who tend to vary more within sessions also tend to vary more across sessions does suggest that there may be a "characterological" aspect to perceptual-motor intraindividual variability that is consistent with a "trait-like" interpretation of the phenomenon. This is supported by the finding (Table 2) that the Sessions 1–3 intraindividual variability correlations tend to be as high as the Sessions 1–2 and Sessions 2–3 correlations. Finally, greater intraindividual variability seems to signal poorer cognitive status in a variety of intellectual tasks. This finding is consistent with an age-related pattern of decline.

Table 4. Correlations of Individual Differences Variables With Between-Session Standard Deviations

Variable	Age	Sex	Educ.	Gc	Gf1	Gf2	Mem.
Tracking error	.22**	.08	−.20**	−.20**	−.37**	−.40**	−.22**
Tracking lag	.22**	.06	−.00	−.08	−.13	−.09	−.22**
Connections							
Same	.37**	−.10	−.13	.03	−.27**	−.31**	−.24**
Alternating	.21*	.04	−.29**	−.35**	−.51**	−.51**	−.37**

Notes: Educ. = education; Gc = composite measure of crystallized intelligence; Gf1 = composite measure of fluid intelligence; Gf2 = a different composite measure of fluid intelligence; Mem. = Measure of specific memory.

$^*p < .05$; $^{**}p < .01$.

Discussion

Our data tell a relatively consistent story that adds significantly to understanding the nature of short-term, intraindividual variability and extends it to perceptual-motor performance. We organize the discussion around the methodological and theoretical issues identified earlier. First, we look at the methodological issue of evaluating the representativeness of a single occasion of measurement.

One of the more compelling findings is that within-person variability is 37% to 53% of the between-person variability when both are expressed in standard deviation units. This substantial range reinforces the idea that any single occasion of measurement might not accurately represent the "typical" performance of an individual. Indeed, it calls into question the very notion of "typical." It challenges the value of even a "working" notion of the classical test theory conception of true score.

There is a pitfall to be avoided in considering the lack of representativeness of single occasion of measurement scores. It can be seen, for example, in the distinction between reliability of measurement and actual lability or variability in the psychological or behavioral quality being measured. An attribute that is highly labile from one occasion to another, such as state anxiety, can still be measured very reliably on any given occasion, so one needs to be careful regarding the reference implicit in such terms as "lack of precision." A given score today may be quite precise but not representative of one's score tomorrow. Thus, if we assume adequate reliability of the measurement operations themselves, the problem arises when a single occasion of measurement does not adequately represent the potential range of scores that characterizes an individual's repertoire.

As was noted earlier, if the within-person score range is very small relative to the between-person score range, intraindividual variability can be ignored without significant consequences. If the converse is true, however, intraindividual variability cannot be ignored and lack of representativeness can be a genuine threat to valid conclusions regarding substantive phenomena. This has implications for both cross-sectional and longitudinal research designs. One of the more serious implications of the lack of representativeness of single-occasion measurements for students of age effects is illustrated by the fact that the average within-person standard deviation corresponds to a cross-sectional age difference of 12 to 27 years, which renders estimates less accurate than they would otherwise be. Moreover, average within-person standard deviations correspond to greater cross-sectional age differences with increased age. Thus, there will be even less precision in the estimates of age effects for older participants.

In longitudinal research, the existence of substantial intraindividual variability can hamper the disentanglement of the effects of aging, effects of prior experience, short-term fluctuation, and measurement error. Yaffe, Browner, Cauley, Launer, and Harris (1999) studied over 8,000 women (average age 70 years) who were retested after 4 to 6 years on the Digit Symbol and Trail Making Part B tests. The average difference was −.145 *SD* on the Trail Making test, and −.163 *SD* on the Digit Symbol test. The estimates of intraindividual variability in the current study (cf. rows 6 and 7 in Table 3) are two to three times those values, thus making it very difficult to detect change at the level of the individual, or to detect correlations among changes in different variables.

Clearly, ignoring short-term variability can lead to confusing true longitudinal change with within-person variability. A design implication that can be drawn from this is the need to more systematically incorporate measurement bursts at every occasion in longitudinal assessment in order to distinguish short-term fluctuation from other influences (Nesselroade, 1991). The measurement burst design would also allow an individual's change to be calibrated relative to his or her own variability instead of between-person variability (cf. Salthouse, Kausler, & Saults, 1986).

The theoretical concern around which this article was constructed relates to the nature of age differences in moment-to-moment, or day-to-day, intraindividual variability. To what extent do these data support the suggestion by Rowe and Kahn (1987) that increased intraindividual variability should be considered a risk factor for mortality in the elderly population? This, too, is a more complicated issue than it may first appear. Age was associated with increased variability, both within and between sessions, and increased intraindividual variability was negatively associated with performance on a battery of cognitive measures. Although the present study does not address the matter directly, the relationship between magnitude of intraindividual variability and mortality may simply represent the normative cross-sectional relationship between age and mortality rather than something having to do directly with increased intraindividual variability.

Finally, an obvious limitation of the current study is that only perceptual-motor variables were studied. However, these findings add to the substantial list of variables that have exhibited notable amounts of intraindividual variability in other studies (e.g., Eizenman et al., 1997; Hertzog et al., 1992; Hultsch, Hertzog, Dixon, & Small, 1998; Li, Aggen, Nesselroade, & Baltes, 2001; Rabbitt, Osman, Moore, & Stollery, 2001; Shammi, Bosman, & Stuss, 1998). As the list of variable domains evincing substantial intraindividual variability grows, it becomes more desirable to move beyond the descriptive, documentation phase toward the building of predictive relationships and their explanations by means of theory.

Acknowledgments

This work was supported by The Institute for Developmental and Health Research Methodology at The University of Virginia and by National Institute on Aging Grants R01 AG19627 to Timothy A. Salthouse and R01 AG18330-02 to John R. Nesselroade. J. Nesselroade is also grateful for the support of the Max Planck Institute for Human Development, Berlin, Germany.

Address correspondence to John R. Nesselroade, Psychology Department, 102 Gilmer Hall, The University of Virginia, P.O. Box 400400, Charlottesville, VA 22904-4400. E-mail: jrn8z@virginia.edu

References

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

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3–20). Madison, WI: University of Wisconsin Press.

Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave data. *Multivariate Behavioral Research, 37*, 127–160.

Browne, M. W., & Nesselroade, J. R. (in press). Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Psychometrics: A festschrift to Roderick P. McDonald.* Mahwah, NJ: Erlbaum.

Butler, A. C., Hokanson, J. E., & Flynn, H. A. (1994). A comparison of self-esteem lability and low trait self-esteem as vulnerability factors for depression. *Journal of Personality and Social Psychology, 66*, 166–177.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: World Book Co.

Eizenman, D. R., Nesselroade, J. R., Featherman, D. L., & Rowe, J. W. (1997). Intra-individual variability in perceived control in an elderly sample: The MacArthur Successful Aging Studies. *Psychology and Aging, 12*, 489–502.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for the kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin, 52*, 217–250.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. (2003). *Statistical modeling of the individual: Rational and application of multivariate time series analysis.* Unpublished manuscript, University of Amsterdam, Department of Psychology.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain and Language, 42*, 248–269.

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged.* Cambridge, England: Cambridge University Press.

Hultsch, D. F., & MacDonald, S. W. (in press). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L.-G. Nilsson, & L. Blackman (Eds.), *New frontiers in cognitive aging.* New York: Oxford University Press.

Kraemer, H. C., & Korner, A. F. (1976). Statistical alternatives in assessing reliability, consistency, and individual differences for quantitative measures of neonates. *Psychological Bulletin, 83*, 914–921.

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.

McArdle, J. J. (1982). *Structural equation modeling of an individual system: Preliminary results from "a case study in episodic alcoholism."* Unpublished manuscript, University of Denver, Department of Psychology.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analysis with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50*, 181–202.

Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modeling intraindividual variability with repeated measures data.* Mahwah, NJ: Erlbaum.

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

Nesselroade, J. R. (2002). Elaborating the different in differential psychology. *Multivariate Behavioral Research, 37*, 543–561.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223–250). Newbury Park, CA: Sage.

Noll, J. G., & Horn, J. L. (1998). Age differences in processes of fluid and crystallized intelligence. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 263–281). Mahwah, NJ: Erlbaum.

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

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

Rowe, J. W., & Kahn, R. L. (1987). Human aging: Usual and successful. *Science, 237*, 143–149.

Salthouse, T. A., Berish, D. E., & Miles, J. D. (2002). The role of cognitive stimulation on the relations between age and cognitive functioning. *Psychology and Aging, 17*, 548–557.

Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging, 18*, 91–110.

Salthouse, T. A., Kausler, D. H., & Saults, J. S. (1986). Groups versus individuals as the comparison unit in cognitive aging research. *Developmental Neuropsychology, 2*, 363–372.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition, 30*, 572–582.

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

Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, and Cognition, 5*, 1–13.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science, 3*, 1–5.

Thouless, R. H. (1936). Test unreliability and function fluctuation. *British Journal of Psychology, 26*, 325–343.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality, 9*, 609–662.

Woodcock, R. W., & Johnson, M. B. (1990). *WJ-R tests of cognitive ability.* Allen, TX: DLM Teaching Resources.

Woodrow, H. (1932). Quotidian variability. *Psychological Review, 39*, 245–256.

Yaffe, K., Browner, W., Cauley, J., Launer, L., & Harris, T. (1999). Association between bone mineral density and cognitive decline in older women. *Journal of the American Geriatrics Society, 47*, 1176–1182.

Received August 25, 2003
Accepted November 11, 2003
Decision Editor: Margie E. Lachman, PhD