Psychology Science, Volume 49, 2007 (2), p. 132-149

Assessing intraindividual variability in sustained attention: reliability, relation to speed and accuracy, and practice effects

HAGEN C. FLEHMIG[1], MICHAEL STEINBORN[2], ROBERT LANGNER[3], ANJA SCHOLZ[4] & KARL WESTHOFF[4]

Abstract

We investigated the psychometric properties of competing measures of sustained attention. 179 subjects were assessed twice within seven day's time with a test designed to measure sustained attention, or concentration, respectively. In addition to traditional performance indices [i.e., speed (M_{RT}) and accuracy (E%)], we evaluated two intraindividual response time (RT) variability measures: standard deviation (SD_{RT}) and coefficient of variation (CV_{RT}). For the overall test, both indices were reliable. SD_{RT} showed good to acceptable retest reliability for all subtests. For CV_{RT}, retest reliability coefficients ranged from very good to not satisfactory. While the reversed-word recognition test proved highly reliable, the mental calculation test and the arrows test were not sufficiently reliable. CV_{RT} was only slightly correlated but SD_{RT} was highly correlated with M_{RT}. In contrast to substantial practice gains for M_{RT}, SD_{RT} and E%, only CV_{RT} proved to be stable. In conclusion, CV_{RT} appears to be a potential index for assessing performance variability: it is reliable for the overall test, only moderately correlated with speed, and virtually not affected by practice. However, before applying CV_{RT} in practical assessment settings, additional research is required to elucidate the impact of task-specific factors on the reliability of this performance measure.

Key words: concentration; sustained attention; intraindividual variability; coefficient of variation; reliability; response time

[1] Hagen C. Flehmig, Psychologisches Institut II, Technische Universität Dresden, Zellescher Weg 10, 01062 Dresden, Germany, Phone: +49-351-463-34004, email: hagen.flehmig@tu-dresden.de

[2] University of Tübingen

[3] University of Tübingen and RWTH Aachen University

[4] Dresden University of Technology

Trial-to-trial variations of response speed in serial choice response time (RT) tasks are well known and have been extensively documented. These intertrial differences in RT have often been attributed to attentional oscillations – a relationship that had already been proposed by Obersteiner (1879) and Guilford (1927). Another early researcher, Kraepelin (1902), who pioneered in developing a procedure for assessing intraindividual variability in sustained concentration ("work curve"), suggested three main sources of performance fluctuations over time: accumulating fatigue, effort variations, and practice effects. Later on, experimental and assessment research tended to overlook the phenomenon of response speed fluctuations (Surwillo, 1975). In that time, RT variability was mostly treated as mesurement error (Fiske & Rice, 1955). Rather late in the history of RT research, Berkson and Baumeister (1967) asserted that intertrial RT variations constitute no measurement error but a phenomenon with reliable individual differences – a finding which was followed by new research efforts in differential psychology (Jensen, 1992; 1998, pp. 225-228). Up to now, however, no comparable progress has been made in psychometric concentration assessment. This study aimed to lessen that backlog. Using a typical concentration test (with three subtests) exemplarily, we examined retest reliability and correlational structure of two intraindividual RT variability measures: standard deviation (SD_{RT}) and coefficient of variation (CV_{RT}). Furthermore, we studied practice effects on RT variability, speed (mean reaction time, M_{RT}), and accuracy (error percentage, *E%*) due to retesting.

Response Time Variability in Serial Choice RT Tasks

In tasks, commonly applied to assess sustained attention, mean (or median) RT is the usual performance measure, besides the number of correct responses. However, measures of central tendency summarize RT distributions only coarsely, without capturing potentially useful information on intraindividual RT variability (Jensen, 1992; Larson & Alderton, 1990; Rabbitt, Osman, Moore, & Stollery, 2001). Often, RT distributions are asymmetrical: they have a steep slope on the left side which is due to a rather narrow range of very fast responses, and they have an elongated right tail, arising from a substantial amount of more broadly distributed slow responses (Leth-Steensen, Elbaz, & Douglas, 2000; Logan, 1992; Ulrich & Miller, 1993; Wagenmakers, Grasman, & Molenaar, 2005). This distributional asymmetry is due to the fact that there is a physiological limit to maximizing response speed but none to response slowing (Ulrich & Miller, 1993; Ulrich, Miller, & Schröter, in press). Thus, RT variability expresses itself chiefly in responses above mean response time (Larson & Alderton, 1990; Wagenmakers et al., 2005).

Sustained attention has often been studied using self-paced continuous RT tasks, which require individuals to actively maintain performance speed and accuracy over the testing period (Appleton, 1967; Bills, 1943; Kraepelin, 1902; E. S. Robinson & Bills, 1924; Sanders & Hoogenboom, 1970; von Voss, 1899; see also, Westhoff & Kluck, 1984). In the German tradition, these tests are often termed "concentration tests" (Bühner, Mangels, Krumm, & Ziegler, 2005; Schmidt-Atzert, Bühner, & Enders, 2006; Smit & Van der Ven, 1995). Such tests require individuals to engage in repetitive activities such as letter cancellation, detecting differences in simple shapes, or continuously adding digits. In contrast to vigilance or go/nogo tasks (e.g., Ballard, 2001; MacDonald, Hultsch, & Bunce, 2006; Reinvang, 1998; Smid, de Witte, Homminga, & van den Bosch, 2006; Smith, Valentino, & Arruda, 2002), concentration tests are usually designed as serial RT tasks, which require individuals to self-pace their speed and trade it off against accuracy (e.g., Bertelson & Joffe, 1963; Schweizer & Moosbrugger, 2004; Westhoff & Kluck, 1984). Typically, speed and accuracy are used to determine an individual's ability to sustain concentration. The term "concentration" has been conceptualized as the ability to maintain attention (i.e., speed and precision) over relatively long time periods (Geissler, 1909; Peak & Boring, 1926; Van Breukelen, 1989; Westhoff & Kluck, 1984). During the task, individuals need to continuously orient attention and adjust perceptuomotor activity to task demands, thereby preventing distraction and irrelevant activity (Posner & Boies, 1971; Posner, Cohen, Choate, Hockey, & Maylor, 1984).

With prolonged time on task, however, work speed has been observed not only to become slower but also less regular (Sanders, 1998, pp. 401-409; Welford, 1984). For example, von Voss (1899) already observed that with prolonged work on a digit addition task, the frequency of long responses increased whereas there was no change in the fastest responses. Similar observations were made by other researchers which therefore regarded fluctuations in work speed as an essential feature of extended concentrative performance (Bills, 1937; Geissler, 1909; Kraepelin, 1902; Obersteiner, 1879; von Voss, 1899). The issue of "mental blocking" was brought to prominence by Bills (1931; 1935), who identified an increase in "extra-long" responses during prolonged colour naming. In most of the early work, blocks were defined as responses longer than a fixed criterion, mostly as responses longer than twice the mean (Bertelson & Joffe, 1963; Bills, 1931, 1935; Bunce, Warr, & Cochrane, 1993; Fiske & Rice, 1955; Sanders & Hoogenboom, 1970).

However, the question of what causes the characteristic work speed fluctuations is still unresolved (Weissman, Roberts, Visscher, & Woldorff, 2006). Previous investigations into the nature of intraindividual RT variability drew the conclusion that occasionally occurring attentional lapses may cause the slower responses (e.g., Bertelson & Joffe, 1963; Bills, 1937; Hockey, 1986; Sanders, 1998, pp. 420-421). The "attentional lapses," or "mental blocks," were believed to be involuntary resting pauses, enforced by the accumulation of fatigue during the task (Bertelson & Joffe, 1963; Sanders & Hoogenboom, 1970). This notion is also supported by studies showing that mental fatigue, as induced by prolonged task performance, primarily affects the upper end of the intraindividual RT distribution (Fiske & Rice, 1955; Welford, 1984). In addition, it has been suggested that occasionally occurring task-irrelevant cognitions (i.e., distractions) are responsible for at least some of the response time outliers (Jensen, 1992; Smallwood et al., 2004; Ulrich & Miller, 1994, p. 34), particularly when it is required to maintain performance over extended time periods (Stuss, Meiran, Guzman, Lafleche, & Willmer, 1996; Stuss, Murphy, Binns, & Alexander, 2003). Taken together, the literature supports the view that intraindividual RT variability in sustained attention tasks is an empirical phenomenon distinct from other performance characteristics and a useful concept for theorizing on "energetical" issues in speeded performance (Pieters, 1985; Sanders, 1983; Van Breukelen, 1989).

Predictive Value of Intraindividual RT Variability

Energetical issues in performance have been extensively discussed in clinical and individual-differences research (cf. Matthews, Davies, Westermann, & Stammers, 2000, pp. 265-285; Welford, 1984). In their seminal work, Baumeister and Kellas (1968) observed that mentally retarded individuals, in comparison to normals, were capable to sustain performance speed for short but not extended time periods. This finding was later confirmed and generalized to intelligence differences in the normal population: the slowest individual responses (i. e., the most dramatic drops in performance speed) were the best predictor for general intelligence (e.g. Larson & Alderton, 1990). Indeed, several individual-differences variables have been shown to affect RT variability, rather than mean RT or error percentage. For example, Bunce, MacDonald and Hultsch (2004) discovered that younger (M = 25 years) and older (M = 69 years) adults can be dissociated by measures of variability rather than by measures of central tendency. This is supported by other studies reporting that the effects of aging primarily express themselves in higher RT variability rather than higher mean RT (e.g. Friedman, 2003; Hultsch, MacDonald, & Dixon, 2002; Shammi, Bosman, & Stuss, 1998; Uttl, Graf, & Cosentino, 2000).

In neuropsychological research on cognitive deficits following brain damage, increased RT variability has been considered an important index of impaired monitoring of self-generated response speed during concentration tasks (Alexander, Stuss, Shallice, Picton, & Gillingham, 2005; Stuss et al., 1996). Compared to healthy controls, disturbances in intertrial RT consistency have been observed in patients with focal frontal lobe lesions (Stuss et al., 2003), traumatic brain injury (S. J. Segalowitz, Dywan, & Unsal, 1997; Stuss, Pogue, Buckle, & Bondar, 1994; Whyte, Polansky, Fleming, Coslett, & Cavallucci, 1995), and closed head injury (Zahn & Mirsky, 1999). Likewise, neurological conditions like epilepsy and dementia or mild cognitive impairment have been found to be associated with higher RT variability (Burton, Strauss, Hultsch, Moll, & Hunter, 2006; Christensen et al., 2005; Collie, Maruff, & Currie, 2002; Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000).

The phenomenon of increased RT fluctuations has also been observed in mental disorders characterized by deficits in the endogenous control of attention, such as schizophrenia (Schwartz et al., 1989; Zahn et al., 1998). Also, attention deficit/hyperactivity disorder (ADHD) is characterized by increased performance variability: in tasks that require maintaining attention over time, ADHD is associated with more frequent extra-long responses (i. e., mental blocks) but also with more frequent fast impulsive responses (Castellanos et al., 2005; Ridderinkhof, Scheres, Oosterlaan, & Sergeant, 2005). For example, Leth-Steensen, Elbaz & Douglas (2000) showed that boys with ADHD did not differ from healthy controls in the average speed of performance but in speed variability. Specifically, ADHD boys produced an RT distribution with elongated right tail, that is, a higher percentage of blocks. Such findings led several researchers to suggest that increased RT variability might represent an etiologically important characteristic of ADHD (Bellgrove, Hawi, Kirley, Gill, & Robertson, 2005). Some recent studies also found associations between anxiety-related personality traits and the stability of basic cognitive operations. For example, Robinson and Tamir (2005) reported that neuroticism is correlated with variability in simple and choice RT tasks, that is, high-neuroticism subjects were found to be more variable in their performance speed than low-neuroticism subjects.

Apart from stable individual differences, research on performance variations is also concerned with "energetical" variables that affect the current state of the organism (Folkard, 1983; Hockey, 1986). The deleterious effects of such situational variables on attentional state and, in turn, on sustained attention performance may be better reflected by variability measures than by measures of central tendency, as evidenced by studies dealing with the impact of prolonged work and fatigue (Healy, Kole, Buck-Gengler, & Bourne, 2004; Henning, Sauter, Salvendy, & Krieg, 1989), sleep loss (Anderson & Horne, 2006), circadian/diurnal rhythms (Bratzke, Rolke, Ulrich, & Peters, 2007; Monk & Carrier, 1997), or alcohol (Maylor, Rabbitt, James, & Kerr, 1992). In conclusion, measures of RT variability appear to better predict an individual's capability to retain an attentional state optimal for task demands than measures of central tendency. Thus, the above findings indicate that variability measures may be of important diagnostic value. However, it is still an open question to what degree RT variability obtained in different RT tasks reflects similar energetical and/or cognitive processes (cf. Weissman et al., 2006). Further, since variability measures can be derived by different calculation procedures, they might differ in their psychometric properties (Fiske & Rice, 1955; Guilford, 1956, pp. 78-103). Therefore, it is important to examine those measures regarding their suitability for assessing performance variability.

Measures of Intraindividual RT Variability

There are multiple indices that may be computed to examine intraindividual variability of performance (Guilford, 1956, pp. 78-103; Slifkin & Newell, 1998). Most early researchers suggested the mean deviation of response times as a measure of performance (Fiske & Rice, 1955; Peak & Boring, 1926; Spearman, 1927). Often, the standard deviation of response times (SD_{RT}) has been used as an index of performance variability. However, simply computing SD_{RT} is problematic (Hultsch et al., 2002). First, SD_{RT} is highly influenced by the individual's average work speed (M_{RT}), indicating that both measures share a substantial proportion of variance in common (Jensen, 1992). Hence, SD_{RT} might be considered a relatively redundant measure of performance. For example, differences in SD_{RT} between younger and older adults might simply reflect the fact that older adults are on average slower than younger adults (Burton et al., 2006; Shammi et al., 1998). Second, systematic changes over time (e.g., practice effects due to retesting) may be present, which do not only affect M_{RT} but, in a similar way SD_{RT} as well (Logan, 1992; Smit & Van der Ven, 1995; Wagenmakers et al., 2005). Taken together, these problems reflect substantial limitations of the utility of $\mathrm{SD}_{\mathrm{RT}}$ as an index of RT variability.

In response to these problems, a number of techniques have been developed to study individual differences in intertrial RT variability while controlling for possible differences in mean M_{RT}: For instance, linear regression has been used to partial out effects of M_{RT} on interindividual differences in SD_{RT}, yielding the residual standard deviation (Wagenmakers et al., 2005), which is a measure of RT variability entirely independent of M_{RT}. Alternatively, the coefficient of variation (CV_{RT}) has been employed to control for interindividual differences in M_{RT} (N.S. Segalowitz, Poulsen, & Segalowitz, 1999). CV_{RT} is a so-called relative variability measure for which each individual's SD_{RT} is related to the individual's mean response time, yielding an index of variability relative to the individual's overall level of work speed. CV_{RT} is calculated by dividing individual SD_{RT} by individual M_{RT}, multiplied by 100: $CV_{RT} = (SD_{RT} / M_{RT})\ x\ 100$ (Guilford, 1956, p. 101). As a result, a measure is obtained that allows to compare intraindividual RT variability even between of individuals who differ very much in their average work speed.

Research overview

The goal of the present study was to examine whether there is useful information in intraindividual RT variability and how to best extract this information to assess sustained attention. Since performance measures have to conform to several "basic" psychometric quality standards to be useful in research as well as in applied settings, we investigated test-retest reliability, correlational structure, and practice effects of two competing RT variability measures (SD_{RT} and CV_{RT}) in a sample of normal individuals.

Reliability. Retest reliability provides information about the consistency of individual test scores in a series of measurements. Usually, it is indexed by the correlation between two measurements of the same test. The reliability coefficient tells us to what extent the test variance is due to "true" individual differences rather than sampling error (Cronbach, 1975, p. 126). If intraindividual RT variability in sustained concentration is a function of subject factors, then we would expect to observe relatively stable individual differences. To determine retest reliability as a psychometric property, it is required to use normal adults as a reference sample, because they are not assumed to be inconsistent per se (Cronbach, 1975, pp. 126-136). Non-normal populations like older adults or neurologically impaired patients would not allow to determine psychometric retest reliability, since these populations are assumed to be largely inconsistent over several testing occasions (Bunce et al., 2004; Burton et al., 2006; Hultsch et al., 2002). Accordingly, we examined test-retest reliability of SD_{RT} and CV_{RT} in neurologically normal adults after a retest interval of one week.

Correlational structure. Since concentration tests are made of uniform and repetitive choice RT tasks, several dependent measures can be computed to determine performance (Smit & Van der Ven, 1995; Westhoff & Kluck, 1984). To justify an index of variability besides the traditional indices of speed and accuracy, it should reflect distinct aspects of concentration ability. Thus, if there were substantial positive correlations between variability (SD_{RT}, CV_{RT}) and speed (M_{RT}), the variability dimension would be redundant (Larson & Alderton, 1990). On the other hand, if there were no more than only small correlations between them (given sufficient reliability), it would suggest considering performance variability as a self-sufficient behavioral expression of sustained attention. Accordingly, correlations between SD_{RT}, CV_{RT}, M_{RT}, and $E\%$ were examined.

Practice Effects. Test scores of sustained attention performance are regarded unstable when behavioral patterns are acquired, so that the to-be-measured ability is confounded by learning effects (Appleton, 1967; Smit & Van der Ven, 1995; Van Breukelen, 1989). Thus, the robustness of performance indices in the face of practice is an important requirement concerning test validity (Falleti, Maruff, Collie, & Darby, 2006; Feinstein, Brown, & Ron, 1994), especially in applied contexts where the amount of prior test experience often cannot be established (Westhoff & Kluck, 1984). Accordingly, effects of practice due to retesting within a one-week interval were examined for SD_{RT}, CV_{RT}, M_{RT}, and $E\%$. It has been observed that practice-related performance gains in various choice RT tasks are equivalent for SD_{RT} and M_{RT} (Logan, 1992; Wagenmakers et al., 2005). Because most of the available findings are based on group-level data, it might be interesting to examine if this prediction also holds for individual differences. Since CV_{RT} is a relative measure of variability, gains due to retesting are expected to be smaller for CV_{RT} than for SD_{RT} (N.S. Segalowitz et al., 1999; Smit & Van der Ven, 1995).

Method

Participants

The data of 179 participants (110 female), aged between 18 and 45 years ($M = 28.0$; $SD = 8.2$), entered the analyses. All subjects had normal or corrected-to-normal vision, and all of them reported to be in good health. The majority of the participants (77 %) reported to have high school graduation, 23 % reported to have secondary school graduation. The sample was recruited via advertisements in a local newspaper and on the university campus.

Material

We used a computerized version of the Complex Concentration Test (CCT, Westhoff & Graubner, 2003). The CCT provides three serial choice RT tasks in figural, numerical and verbal modalities to assess sustained attention performance. The CCT consists of three subtests presented in the following order: arrows test (figural stimuli), mental calculation test (numerical stimuli), and reversed-word recognition test (verbal stimuli). Each subtest requires self-paced serial responding to targets among distractors, which is to be done as fast and accurately as possible. Each test-item is presented until the subject's response and is followed immediately afterwards by the next item. For each subtest, task complexity is varied across five levels. Each subtest has a test duration of 10 min, amounting to an overall test duration of 30 min. Responses were recorded by a conventional computer keyboard with color-coded shift-keys (left: red; right: green), connected to an IBM-compatible computer.

Arrows subtest. The stimuli consist of four different types of arrows pointing into one of four different directions. Different arrows are randomly presented one after another in a line. Subjects are instructed to respond to targets (i.e., simple arrows pointing to the upper right corner) by pressing the right shift-key, and to non-targets (any other combination of arrow type and direction) by pressing the left shift-key. Task complexity is achieved by using arrows of different figural complexity and dimensional overlap. The relation of targets (75 %) to distractors (25 %) is constant.

Mental calculation subtest. In this subtest, the subject is presented with simple chained addition and subtraction tasks and a possible result. The calculation chains consist of two to four positive integers ranging from 1 to 40, leading to a result between 1 and 40. Participants have to judge the correctness of the presented result by pressing the right shift-key in response to a correct result and the left shift-key in response to an incorrect one. Complexity is enhanced by increasing chain length, that is, by adding a summand or subtrahend to the calculation term. Additional variation in complexity is realized by using not only pure addition tasks but also mixed addition-subtraction terms. The proportion of correct trials remains 75 % throughout.

Reversed-word recognition subtest. In this subtest, a regularly used German-language noun is simultaneously presented with a nonsensical letter sequence. Subjects are instructed to check whether this letter sequence constitutes the exact reversal of the noun. If so, participants have to respond as fast as possible by pressing the right shift-key, otherwise by pressing the left shift-key. Inexact reversals have been derived from exact reversals by randomly switching letter positions. Complexity is varied by using words of different length, ranging from four to eight letters. The proportion of exact word reversals remains at a 50 % level throughout.

For the overall CCT, M_{RT} and SD_{RT} are obtained by averaging the respective z-transformed values of the three subtests. Overall accuracy (*E%*) and CV_{RT} are obtained by averaging the raw scores of the three subtests.

Procedure

The CCT was administered twice, with an intertest interval of seven days. The procedure at each test session was exactly the same. After a short instruction, a warm-up session was done. This was followed by the three subtests presented in the following order: (1) arrows test, (2) mental calculation test, (3) reversed-word recognition test. The testing was done in a noise-shielded room, in which participants were seated about 80 cm in front of a computer screen.

Results

Nonparametric statistics were used whenever appropriate, since Kolmogorov–Smirnov tests indicated that most performance indices were not normally distributed. We applied the multitrait-multimethod (MTMM) procedure (Campbell & Fiske, 1959) as a heuristic aid to

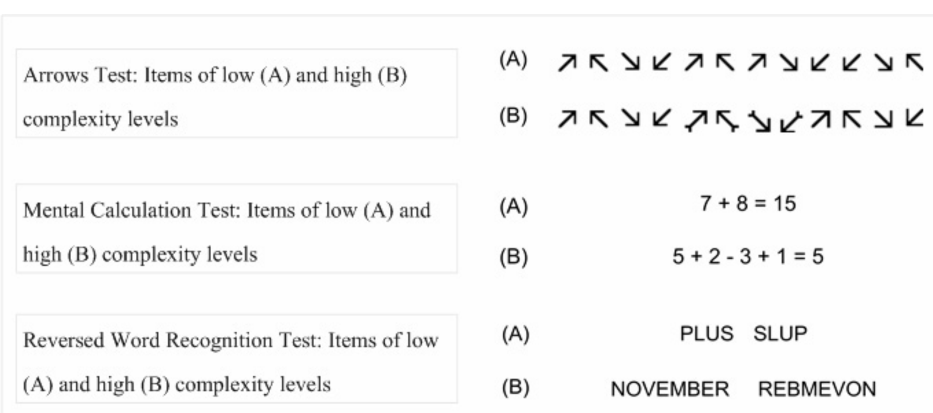

Figure 1:
Examples of items of the CCT-subtests: arrows test, mental calculation test, and reversed-word recognition test. Items are displayed for low and high complexity.

Table 1:
Multitrait-Multimethod Matrix. Correlations between Speed, Variability, and Accuracy Among Subtests and Total Scores of the Complex Concentration Test at the Two Testing Sessions

N = 179		O (Overall Test)				A (Arrows Subtest)				C (Calculation Subtest)				R (Reversed-Word Subtest)			
		M_{RT}	SD_{RT}	CV_{RT}	$E\%$	M_{RT}	SD_{RT}	CV_{RT}	$E\%$	M_{RT}	SD_{RT}	CV_{RT}	$E\%$	M_{RT}	SD_{RT}	CV_{RT}	$E\%$
O	M_{RT}	.91	.84	.24		.57	.56	.36	.20	.52	.47			.67	.58		.22
	SD_{RT}	.79	.84	.60	.21	.48	.59	.50		.54	.55			.62	.65	.24	
	CV_{RT}		.58	.80	.22			.28		.34	.36	.23		.29	.42	.21	
	$E\%$		.20		.71				.61	.25	.23		.52				.56
A	M_{RT}	.53	.34			.82	.80	.35		.42	.34			.57	.47		
	SD_{RT}	.58	.49			.73	.72	.82		.44	.44			.55	.56		
	CV_{RT}	.39	.44			.28	.82	.60		.31	.38	.25		.32	.45	.20	
	$E\%$				.54				.60				.51				.53
C	M_{RT}	.41	.56	.41	.21	.29	.42	.36	.23	.89	.88		.26	.53	.53	.27	.22
	SD_{RT}	.31	.55	.45			.34	.34		.90	.85	.54		.50	.52	.24	
	CV_{RT}			.27						.21	.57	.62					
	$E\%$				.47				.47	.28	.21		.62				.45
R	M_{RT}	.64	.51	.21		.61	.55	.30		.43	.36			.89	.68		
	SD_{RT}	.56	.62	.41		.39	.48	.40		.51	.52	.21		.67	.76	.36	
	CV_{RT}		.30	.29	.20					.32	.36	.26			.33	.88	
	$E\%$				.50				.44				.42				.69

Notes: Correlation coefficients (Rho) for the first session is shown above, for the second session below the main diagonal. Test-retest reliability is shown on the main diagonal (denoted with dark grey). Discriminant validity is denoted with medium grey. Convergent validity is denoted with light grey. Correlations smaller than .20 ($p > .01$) are not shown.

systematically analyze the relationships between the indices of concentration performance, that is, speed (M_{RT}), variability (SD_{RT} and CV_{RT}), and accuracy (*E%*). The resulting matrix presents Spearman correlations between all performance indices of each subtest and overall test at both test sessions. Results are shown in Table 1.

Retest reliability. Reliability coefficients are shown along the main diagonal of the correlation matrix, presenting the correlations between the first and the second test administration. As expected, M_{RT} was highly reliable for each of the subtests and the overall test (≥ .82). SD_{RT} showed good reliability for the overall (.85) and reversed-word recognition test (.76), as well as for the mental calculation test (.85). For the arrows tests, however, SD_{RT} failed the reliability criterion. In contrast, the measure of relative variability, CV_{RT}, showed good reliability for the overall test (.80) and very good reliability for the reversed-word recognition subtest (.88). Yet, for the arrows and the mental calculation subtests, retest reliability of CV_{RT} was not satisfactory (< .60). Accuracy (*E%*) showed less reliability compared to M_{RT} but was satisfactorily reliable for the overall test and the reversed-word recognition subtest (> .69).

Correlational structure. The correlations between the different performance indices are shown in Table 1. Inspection of the MTMM-matrix indicates significant relationships between SD_{RT} and M_{RT} (.67-.90). In contrast, consistently low correlations are to be found between CV_{RT} and M_{RT} (.09-.35) across all subtests and the overall test. Between the two variability measures (SD_{RT} and CV_{RT}) and *E%*, only negligible correlations were found (< .22). Thus, for CV_{RT}, but not for SD_{RT}, discriminant validity could be demonstrated: SD_{RT} and M_{RT} appear to share substantial variance, whereas the relative measure of variability, CV_{RT}, reflects aspects of sustained performance that are not covered by the traditional indices.

Practice effects. In order to examine stability in the face of repeated testing, we performed a multivariate repeated-measures analysis of variance (MANOVA) including all performance measures. We choose to use a MANOVA instead of nonparametric tests, since it is more sensitive and, at our sample size, sufficiently robust against violations of the normal-distribution assumption. Also, it obviates post-hoc correction for multiple comparisons. The main results are shown in Table 2 and Figure 2.

The statistical analyses revealed significant multivariate effects from the first to the second testing session for all three subtests: arrows subtest: $F(4, 175) = 110.56$, $p < .01$, $\eta^2 = 0.72$; calculation subtest: $F(4, 175) = 18.07$, $p < .01$, $\eta^2 = 0.29$; reversed-word subtest: $F(4, 175) = 87.48$, $p < .01$, $\eta^2 = 0.67$. Planned single comparisons showed significant changes for SD_{RT}, M_{RT}, and *E%* across all subtests: M_{RT} became significantly shorter after practice (9-30 %), also SD_{RT} (7-58 %) and *E%* (12-60 %) decreased significantly. In contrast, relative variability as indexed by CV_{RT} proved to be generally invariant to the effects of practice. Specifically, statistically significant but practically negligible changes were observed for the arrows subtest [19 %, $F(1, 178) = 17.05$, $p < .01$, $\eta^2 = .09$]; no significant reductions were observed in the reversed-word recognition [1 %, $F(1, 178) = 0.14$, $p = .71$, $\eta^2 = .00$] and the mental calculation [2 %, $F(1, 178) = 3.2$, $p = .07$, $\eta^2 = .02$] subtests. Consequently, CV_{RT} proved to be virtually unaffected by practice effects due to repeated testing.

Table 2:
Performance Changes in Speed, Variability, and Accuracy in the Two Testing Session.

		Arrows Subtest			Calculation Subtest			Reversed-Word Subtest		
	Session	*M*	Gains (%)	η^2	*M*	Gains (%)	η^2	*M*	Gains (%)	η^2
M_{RT}	1	567			3525			2339		
	2	435	30**	.65	3222	9**	.26	1936	21**	.48
SD_{RT}	1	431			3271			1735		
	2	273	58**	.35	3069	7**	.06	1451	20**	.13
CV_{RT}	1	73			92			79		
	2	62	19**	.09	94	2	.02	80	1	.00
E%	1	3.1			4.5			7.9		
	2	1.9	60**	.29	4.1	12*	.03	5.2	52**	.38

Notes: * $p < .05$; ** $p < .01$; η^2 = effect size; M_{RT} = mean response time; SD_{RT} = standard deviation of response times; CV_{RT} = coefficient of variation of response times; *E*% = error percentage.

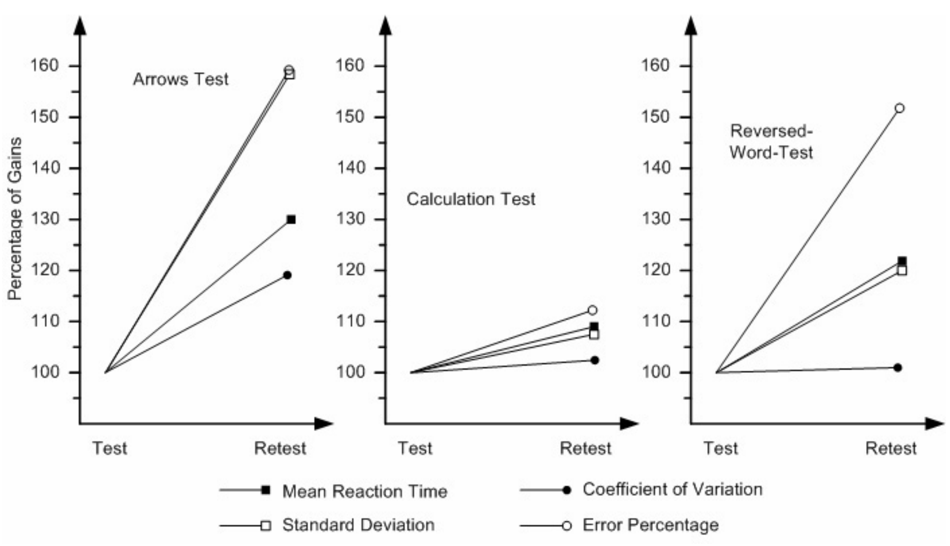

Figure 2:
Effects of repeated testing for the measures of speed, accuracy, and variability. Note that percentage gains indicate reductions in absolute values. M_{RT} = average response time; SD_{RT} = standard deviation of response times; CV_{RT} = coefficient of variation of response times; *E%* = error percentage.

Discussion

The present study evaluated the psychometric properties of two competing measures of intraindividual RT variability (SD_{RT} and CV_{RT}) in sustained attention performance. Exemplarily, we used the Complex Concentration Test (CCT, Westhoff & Graubner, 2003) to assess sustained attention. We studied retest reliability, correlational structure (i.e., interrelationships among different performance indices) and practice effects of four measures of concentration performance (M_{RT}, *E%*, SD_{RT}, and CV_{RT}). For the overall test score of the CCT, reliable interindividual differences in RT variability could be observed using SD_{RT} or CV_{RT}. In contrast to SD_{RT}, CV_{RT} was shown to be only slightly correlated with M_{RT} and *E%*, the traditional measures of sustained attention. Finally, our analyses revealed that CV_{RT} was less affected by practice compared to SD_{RT} and the other indices (M_{RT} and *E%*), which showed substantial gains at the second test administration.

Reliability. SD_{RT} showed good to acceptable retest reliability. For CV_{RT}, retest reliability coefficients ranged from very good to not satisfactory. While the reversed-word recognition test proved highly reliable, the mental calculation test and the arrows test were not sufficiently reliable. The reason for these different reliabilities may be a differential impact of changes in attentional state on the subtests. That is to say, the different concentration tasks included in the test may be differentially vulnerable to changes in energetical factors (e.g., Rabbitt et al., 2001). However, the psychometric properties of the mental calculation and the arrows subtests might not be appropriate to assess fluctuations in performance. Originally, the CCT was not designed to measure performance variability. Therefore, no specific analyses of item properties had been done to fit such demands. The high reliability of the reversed-word recognition test might potentially benefit from its relative position as the final subtest in the CCT. That is to say, fatigue accumulating over the preceding 20-min test duration might serve as an additional factor of variance that accentuates individual differences in performance variability such that they can be reliably detected. Pronounced fatigue effects in repetitive tasks have even been reported after relatively short time periods of about 20-30 min, but not (or to a lesser degree) in test batteries that include a variety of tasks (Matthews et al., 2000, pp. 207-212; Uttl et al., 2000). Most important, the effects of prolonged work have been found to be distinctively reflected by RT variability rather than mean RT or error percentage (Sanders, 1998, pp. 403-408; Welford, 1984). Thus, it seems promising for the future to examine the utility of CV_{RT} as an index of fatigue, satiation, or exhaustion. In conclusion, CV_{RT} qualified as a reliable measure of performance fluctuations for the overall test. However, care must be taken when considering CV_{RT} for the interpretation of performance in single subtests. Nevertheless, because the overall test is of primary importance in practical assessment contexts, CV_{RT} may serve as a reliable measure of additional aspects of sustained attention.

Correlational structure. Relative variability measures (i.e., CV_{RT}) are assumed to reflect behavioral aspects of performance which are not yet included in traditional concentration tests (Smit & Van der Ven, 1995). As expected, only slight correlations were found between CV_{RT} and M_{RT}, and virtually no correlations were found between CV_{RT} and *E%*. SD_{RT}, however, was found to be highly correlated with M_{RT}, as was expected according to earlier findings (Jensen, 1992; Larson & Alderton, 1990). Thus, in contrast to SD_{RT}, CV_{RT} was shown to have discriminant validity and therefore can be taken to reflect a self-sufficient behavioral dimension of task performance. However, for reasons of reliability, this claim must be restricted here to CV_{RT} of the overall and reversed-word recognition test. Further investigations should deal with the predictive validity of CV_{RT} to clarify what aspect of performance is precisely reflected by this relative measure of RT variability. However, the present study did not concern predictive validity but focused on the inter-correlational structure of different performance measures. Intuitively, relative RT variability seems to reflect distractibility (Smit & Van der Ven, 1995). This has also been suggested by other authors (Leth-Steensen et al., 2000; Wagenmakers et al., 2005; West, Murphy, Armilio, Craik, & Stuss, 2002; Westhoff & Kluck, 1984), who view RT variability as being primarily caused by occasional very slow responses due to attentional lapses.

Practice effects. CV_{RT}, but not SD_{RT}, can be considered invariant to practice effects arising from retesting after one week. While all the other indices changed substantially across the two testing sessions, virtually no changes occurred for CV_{RT}. This raises the question of why a measure of intraindividual RT variability should, in principle, not be susceptible to practice effects. Specifically, why should speed improvements with practice not become more consistent, too? (Feinstein et al., 1994; Logan, 1992; Rabbitt & Banerij, 1989). This has been shown for the relation between M_{RT} and SD_{RT}, but appears not to apply to CV_{RT} (Logan, 1992; N. S. Segalowitz & Segalowitz, 1993). This feature makes CV_{RT} quite interesting for practical applications, in which test validity is often compromised by the effects of prior test experiences. In "real-life" assessment situations, such as neuropsychological rehabilitation, occupational aptitude testing, or school psychology, retesting is fairly common (Feinstein et al., 1994; Westhoff & Kluck, 1984). Thus, if a performance measure is known to be significantly affected by practice, and an individual's performance level before practice cannot be established, it becomes difficult to separate potential practice effects from the individual's ability, which the test was designed to measure (Cronbach, 1975, pp. 310-312). Failure to use appropriate control techniques would then lead to erroneous inferences about the aptitude of the individual tested. Especially in the field of achievement testing, CV_{RT} might therefore turn out to be a useful index of performance, since it is not – or only to a minor degree – affected by repeated testing. Of course, further research is needed to examine whether invariance to practice effects is a general property of CV_{RT} or only specific to the tasks reported in the present study.

Conclusions. The coefficient of variation has been shown to be a reliable measure of intraindividual RT variability with regard to the overall performance in the CCT. It appears to reflect aspects of task performance that are not captured by traditional performance measures. A further intriguing feature is its invariance to practice (in contrast to SD_{RT}), at least to retesting. According to our findings, CV_{RT} might be a potential candidate for characterizing additional aspects of sustained attention performance, in research and applied contexts. However, before applying CV_{RT} in practical assessment settings, additional research is required to elucidate the impact of task-specific factors on the reliability of this performance measure. Moreover, in the absence of external criteria across different domains, it might be difficult and premature to even tentatively decide on one measure. Thus, future research efforts should also be directed at further elucidating the predictive value of the relative variability of performance as indicated by CV_{RT}. For reasons of generalizability, research also needs to examine the psychometric properties of CV_{RT} in different serial and discrete choice RT tasks. To this end, it might be beneficial to use CV_{RT} in clinical research in populations with concentration deficits and to manipulate some presumably important situational variables, such as fatigue, motivation, or stress, which are thought to influence sustained attention/concentration task performance (Hockey, 1986; Matthews et al., 2000, chap. 12 and 14). In case of establishing solid relationships between such conditions or variables influencing attentional state and relative performance variability, it is hoped to gain further insights into the complex interplay between “energetics” and “information processing” (Sanders, 1983) in normal as well as pathological functioning.

References

Alexander, M. P., Stuss, D. T., Shallice, T., Picton, T. W., & Gillingham, S. (2005). Impaired concentration due to frontal lobe damage from two distinct lesion sites. *Neurology, 65*, 572-579.

Anderson, C., & Horne, J. A. (2006). Sleepiness enhances distraction during a monotonous task. *Sleep, 29*, 573-576.

Appleton, W. S. (1967). Concentration. The phenomenon and its disruption. *Archives of General Psychiatry, 16*, 373-381.

Ballard, J. C. (2001). Assessing attention: Comparison of response-inhibition and traditional continuous performance tests. *Journal of Clinical and Experimental Neuropsychology, 23*, 331-350.

Baumeister, A. A., & Kellas, G. (1968). Distribution of reaction times of retardates and normals. *American Journal of Mental Deficiency, 72*, 715-718.

Bellgrove, M. A., Hawi, Z., Kirley, A., Gill, M., & Robertson, I. H. (2005). Dissecting the attention-deficit/hyperactivity disorder (ADHD) phenotype: Sustained attention, response variability and spatial attentional asymmetries in relation to dopamine transporter (DAT1) genotype. *Neuropsychologia, 43*, 1847-1857.

Berkson, G., & Baumeister, A. (1967). Reaction time variability of mental defectives and normals. *American Journal of Mental Deficiency, 72*, 262-266.

Bertelson, P., & Joffe, R. (1963). Blockings in prolonged serial responding. *Ergonomics, 6*, 109-116.

Bills, A. G. (1931). Blocking: A new principle of mental fatigue. *American Journal of Psychology, 43*, 230-245.

Bills, A. G. (1935). Some causal factors in mental blocking. *Journal of Experimental Psychology, 18*, 172-185.

Bills, A. G. (1937). Facilitation and inhibition in mental work. *Psychological Bulletin, 34*, 286-309.

Bills, A. G. (1943). *The psychology of efficiency. A discussion of the hygiene of mental work.* New York: Harper & Brothers.

Bratzke, D., Rolke, B., Ulrich, R., & Peters, M. (2007). Central slowing during the night. *Psychological Science, 18*, 456-461.

Bühner, M., Mangels, M., Krumm, S., & Ziegler, M. (2005). Are working memory and attention related constructs? *Journal of Individual Differences, 26*, 121-131.

Bunce, D. J., MacDonald, S. W., & Hultsch, D. F. (2004). Inconsistency in serial choice decision and motor reaction times dissociate in younger and older adults. *Brain and Cognition, 56*, 320-327.

Bunce, D. J., Warr, P. B., & Cochrane, T. (1993). Blocks in choice responding as a function of age and physical fitness. *Psychology and Aging, 8*, 26-33.

Burton, C. L., Strauss, E., Hultsch, D. F., Moll, A., & Hunter, M. A. (2006). Intraindividual variability as a marker of neurological dysfunction: A comparison of Alzheimer's disease and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology, 28*, 67-83.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Castellanos, F. X., Sonuga-Barke, E. J. S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intraindividual variability. *Biological Psychiatry, 57*, 1416-1423.

Christensen, H., Dear, K. B. G., Anstey, K. J., Parslow, R. A., Sachdev, P., & Jorm, A. F. (2005). Within-occasion intraindividual variability and preclinical diagnostic status: Is intraindividual variability an indicator of mild cognitive impairment? *Neuropsychology, 19*, 309-317.

Collie, A., Maruff, P., & Currie, J. (2002). Behavioral characterization of mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 24*, 720-733.

Cronbach, L. (1975). *Essentials of psychological testing*. New York: Harper & Row.

Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology, 28*, 1095-1112.

Feinstein, A., Brown, R., & Ron, M. (1994). Effects of practice of serial tests of attention in healthy subjects. *Journal of Clinical and Experimental Neuropsychology, 16*, 436-447.

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

Folkard, S. (1983). Diurnal variations in human performance. In G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 245-269). Chichester: John Wiley.

Friedman, D. (2003). Cognition and aging: A highly selective overview of event-related potential (ERP) data. *Journal of Clinical and Experimental Neuropsychology, 25*, 702-720.

Geissler, L. R. (1909). The measurement of attention. *American Journal of Psychology, 20*, 473-529.

Guilford, J. P. (1927). Fluctuations of attention with weak visual stimuli. *American Journal of Psychology, 38*, 534-583.

Guilford, J. P. (1956). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10,* 188-199.

Henning, R. A., Sauter, S. L., Salvendy, G., & Krieg, E. F., Jr. (1989). Microbreak length, performance, and stress in a data entry task. *Ergonomics, 32*, 855-864.

Hockey, G. R. J. (1986). Operator efficiency as a function of effects of environmental stress, fatigue, and circadian rhythm. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. chap. 44). New York: Wiley.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological and Social Sciences, 57*, 101-115.

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.

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences, 13*, 869-881.

Jensen, A. R. (1998). *The g-factor: The science of mental ability*. New York: Praeger.
Kraepelin, E. (1902). Die Arbeitskurve [The work curve]. *Philosophische Studien, 19*, 459-507.
Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14*, 309-325.
Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica, 104*, 167-190.
Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 883-914.
MacDonald, S. W., Hultsch, D. F., & Bunce, D. (2006). Intraindividual variability in vigilance performance: Does degrading visual stimuli mimic age-related "neural noise"? *Journal of Clinical and Experimental Neuropsychology, 28*, 655-675.
Matthews, G., Davies, D. R., Westermann, S. J., & Stammers, R. B. (2000). *Human performance: Cognition, stress, and individual differences*. Hove, East Sussex: Psychology Press.
Maylor, E., Rabbitt, P. M. A., James, G. H., & Kerr, S. A. (1992). Effects of alcohol, practice and task complexity on reaction time distributions. *Quarterly Journal of Experimental Psychology, 44(A)*, 119-139.
Monk, T. H., & Carrier, J. (1997). Speed of mental processing in the middle of the night. *Sleep, 20*, 399-401.
Obersteiner, H. (1879). Experimental researches on attention. *Brain, 1*, 439-453.
Peak, H., & Boring, E. G. (1926). The factor of speed in intelligence. *Journal of Experimental Psychology, 9*, 71-94.
Pieters, J. P. M. (1985). Reaction time analysis of simple mental tasks: A general approach. *Acta Psychologica, 59*, 227-269.
Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78*, 391-408.
Posner, M. I., Cohen, Y., Choate, L. S., Hockey, G. R. J., & Maylor, E. (1984). Sustained concentration: Passive filtering or active orienting. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 49-65). Hillsdale, NY: Lawrence Erlbaum.
Rabbitt, P. M. A., & Banerij, N. (1989). How does very prolonged practice improve decision speed? *Journal of Experimental Psychology: General, 118*, 338-345.
Rabbitt, P. M. A., 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, 54(A)*, 981-1003.
Reinvang, I. (1998). Validation of reaction time in continuous performance tasks as an index of attention by electrophysiological measures. *Journal of Clinical and Experimental Neuropsychology, 20*, 885-897.
Ridderinkhof, K. R., Scheres, A., Oosterlaan, J., & Sergeant, J. A. (2005). Delta plots in the study of individual differences: New tools reveal response inhibition deficits in AD/HD that are eliminated by methylphenidate treatment. *Journal of Abnormal Psychology, 114*, 197-215.
Robinson, E. S., & Bills, A. G. (1924). Two factors in the work decrement. *Journal of Experimental Psychology, 9*, 415-443.
Robinson, M. D., & Tamir, M. (2005). Neuroticism as mental noise: A relation between neuroticism and reaction time standard deviations. *Journal of Personality and Social Psychology, 89*, 107-114.
Sanders, A. F. (1983). Toward a model of stress and human performance. *Acta Psychologica, 53*, 61-97.

Sanders, A. F. (1998). *Elements of human performance*. Mahwah, NJ: Lawrence Erlbaum.
Sanders, A. F., & Hoogenboom, W. (1970). On the effects of continuous active work on performance. *Acta Psychologica, 33*, 414-431.
Schmidt-Atzert, L., Bühner, M., & Enders, P. (2006). Messen Konzentrationstests Konzentration? Eine Analyse von Konzentrationstestleistungen [Does performance in concentration tests reflect the ability to concentrate? An analysis of performance in various concentration tests]. *Diagnostica, 52*, 33-44.
Schwartz, F., Carr, A. C., Munich, R. L., Glauber, S., Lesser, B., & Murray, J. (1989). Reaction time impairment in schizophrenia and affective illness: The role of attention. *Biological Psychiatry, 25*, 540-548.
Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence, 32*, 329-347.
Segalowitz, N. S., Poulsen, C., & Segalowitz, S. J. (1999). RT coefficient of variation is differentially sensitive to executive control involvement in an attention switching task. *Brain and Cognition, 40*, 255-258.
Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369-385.
Segalowitz, S. J., Dywan, J., & Unsal, A. (1997). Attentional factors in response time variability after traumatic brain injury: An ERP study. *Journal of the International Neuropsychological Society, 3*, 95-107.
Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, and Cognition, 5*, 1-13.
Slifkin, A. B., & Newell, K. M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science, 7*, 170-177.
Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., et al. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13*, 657-690.
Smid, H. G., de Witte, M. R., Homminga, I., & van den Bosch, R. J. (2006). Sustained and transient attention in the continuous performance task. *Journal of Clinical and Experimental Neuropsychology, 28*, 859-883.
Smit, J. C., & Van der Ven, A. G. H. S. (1995). Inhibition and speed in concentration tests: The Poisson Inhibition Model. *Journal of Mathematical Psychology, 39*, 265-274.
Smith, K. J., Valentino, D. A., & Arruda, J. E. (2002). Measures of variations in performance during a sustained attention task. *Journal of Clinical and Experimental Neuropsychology, 24*, 828-839.
Spearman, C. (1927). *The abilities of man*. New York: MacMillan.
Stuss, D. T., Meiran, N., Guzman, A., Lafleche, G., & Willmer, J. (1996). Do long tests yield a more accurate diagnosis of dementia than short tests? A comparison of 5 neuropsychological tests. *Archives of Neurology, 53*, 1033-1039.
Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: the frontal lobes control individual performance variability. *Brain, 126*, 2363-2380.
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.
Surwillo, W. W. (1975). Reaction time variability, periodicities in reaction time distributions, and the EEG gating signal hypothesis. *Biological Psychology, 3*, 247-261.

Ulrich, R., & Miller, J. O. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology, 37*, 513-525.
Ulrich, R., & Miller, J. O. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*, 34-80.
Ulrich, R., Miller, J. O., & Schröter, H. (in press). Testing the race model inequality: An algorithm and computer programms. *Behavior Research Methods*.
Uttl, B., Graf, P., & Cosentino, S. (2000). Exacting assessments: do older adults fatigue more quickly? *Journal of Clinical and Experimental Neuropsychology, 22*, 496-507.
Van Breukelen, G. J. P. (1989). Concentration, speed and precision in simple mental tasks. In E. E. C. I. Roskam & R. Suck (Eds.), *Progress in Mathematical Psychology* (Vol. 1, pp. 175-193). Amsterdam: Elsevier.
von Voss, G. (1899). Über Schwankungen der geistigen Arbeitsleistung [On the variance of mental performance]. *Psychologische Arbeiten, 2*, 399-449.
Wagenmakers, E. J., Grasman, R. P. P. P., & Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology, 49*, 195-204.
Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*, 971-978.
Welford, A. T. (1984). Relationships between reaction time and fatigue, stress, age, and sex. In A. T. Welford (Ed.), *Reaction times* (pp. 321-354). London: Academic Press.
West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition, 49*, 402-419.
Westhoff, K., & Graubner, J. (2003). Konstruktion eines komplexen Konzentrationstests [Construction of a Complex Concentration Test]. *Diagnostica, 49*, 110-119.
Westhoff, K., & Kluck, M. L. (1984). Ansätze einer Theorie konzentrativer Leistungen [Towards a theory of concentration performance]. *Diagnostica, 30*, 167-183.
Whyte, J., Polansky, M., Fleming, M., Coslett, H. B., & Cavallucci, C. (1995). Sustained arousal and attention after traumatic brain injury. *Neuropsychologia, 33*, 797-813.
Zahn, T. P., Jacobsen, L. K., Gordon, C. T., McKenna, K., Frazier, J. A., & Rapoport, J. L. (1998). Attention deficits in childhood-onset schizophrenia: Reaction time studies. *Journal of Abnormal Psychology, 107*, 97-108.
Zahn, T. P., & Mirsky, A. F. (1999). Reaction time indicators of attention deficits in closed head injury. *Journal of Clinical and Experimental Neuropsychology, 21*, 352-367.

Author Note

Parts of this study were presented at the 44th conference of the Deutsche Gesellschaft für Psychologie, Göttingen, 2004. We thank Isabelle Nestler, Franziska Klebert, Olja Lengefeld, Jana Grischkowski and Anka Langer for help with collecting the data, and two anonymous reviewers for helpful comments on an earlier version of this paper.