Pergamon

Archives of Clinical Neuropsychology
17 (2002) 547–565

Archives
of
CLINICAL
NEUROPSYCHOLOGY

The Adaptive Digit Ordering Test
Clinical application, reliability, and validity of a verbal working memory test

Katja Werheid[a,*], Christian Hoppe[b], Angelika Thöne[a], Ulrich Müller[c], Martina Müngersdorf[d], D. Yves von Cramon[a,e]

[a] Department of Cognitive Neurology, University of Leipzig, Liebigstrasse 22a, 04105 Leipzig, Germany
[b] University Hospital of Epileptology, Bonn, Germany
[c] Department of Psychiatry, University of Leipzig, Leipzig, Germany
[d] Clinic of Neurology, University Hospital Carl Gustav Carus, Dresden, Germany
[e] Max-Planck-Institute of Cognitive Neuroscience, Leipzig, Germany

Accepted 20 April 2001

Abstract

In the face of ample experimental evidence on the importance of working memory capacity for everyday life, there is a growing need for measures suited for clinical assessment of working memory. For this purpose, the Adaptive Digit Ordering Test (DOT-A), a new version of a digit ordering test introduced by Cooper et al. [Brain 114 (1991) 2095], was developed in analogy to the Digit spans. In Study 1, we investigated DOT-A performance in patients with Parkinson's disease (PD) and patients with frontal lobe damage, as these groups often exhibit working memory impairments within the framework of executive dysfunctions. In comparison with matched controls, both patient groups showed reduced performance in DOT-A but not in Digit span performance. This pattern was found to be particularly sensitive for patients with PD. In Study 2, DOT-A performance was assessed in 50 healthy subjects carefully selected according to demographic criteria in order to ensure representativity. Parallel test and split-half correlations indicated sufficient reliability of the DOT-A. Concurrent validity was confirmed by significant correlation with a well-established working memory test (two-back task). We conclude that DOT-A is a promising diagnostic instrument, and its economy and direct comparability to the

---

* Corresponding author. Tel.: +49-341-99-40-265; fax: +49-341-99-40-204.
E-mail address: werheid@cns.mpg.de (K. Werheid).

PII: S0887-6177(01)00134-2

Wechsler Digit spans and high sensitivity for patient populations make it especially well-suited for assessment in clinical practice.

Keywords: Working memory; Neuropsychological tests; Frontal lobe; Brain damage, chronic; Parkinson disease; Wechsler scales

---

1. Introduction

Working memory has been a major topic of cognitive research during the last 25 years. Since Baddeley and Hitch's (1974) first proposal, a variety of models have been presented, converging in the concept of working memory as processes that are "involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition" (Miyake & Shah, 1999, p. 450).

The basic assumption that working memory involves more than temporary storage of information, i.e., short-term memory, emerged from lesion studies in nonhuman primates (Petrides, 1994), neuroimaging studies (for a review, see Smith & Jonides, 1997, 1999), and behavioral research (e.g., Engle, Tuholski, Laughlin, & Conway, 1999). Within Baddeley's (1986) conceptual framework, this additional feature would be part of a "central executive." Elsewhere, it has been termed "focus of attention" (Cowan, 1995) or "controlled attention" (Conway & Engle, 1994). Recent findings have questioned the concept of a central executive, preferring instead the notion of separable, heterogeneous functions associated with different neural substrates (e.g., Parkin, 1998). Irrespective of these conceptual differences, theorists agree that the additional working memory component (a) is part of the so-called executive functions, (b) is involved in nearly all higher-level cognitive tasks such as language comprehension (King & Just, 1991) or following directions (Engle, Carullo, & Collins, 1991), and (c) depends upon the integrity of the prefrontal cortex. The localization of working memory in prefrontal brain areas has been shown by several recent studies using PET or fMRI (D'Esposito & Postle, 1999; Owen et al., 1999; Postle, Berger, & D'Esposito, 1999), thereby confirming the findings of clinical studies on patients with frontal lobe lesions (FR) and patients with Parkinson's disease (PD). In the following, studies on working memory deficits in patients with FR and in patients with PD are reviewed.

The high prevalence of impaired executive functions in patients with FR has been known since the early days of clinical neuropsychology (Hebb, 1939). Recent research has focused on single aspects of the complex construct of executive functioning and revealed working memory impairments among other deficits in this group. D'Esposito, Postle, Ballard, and Lease (1999) reviewed 11 behavioral studies on short-term memory and working memory in patients with prefrontal cortical lesions and found that none of them reported a significant deficit in short-term retention measured by verbal or spatial span tests. In contrast, patients with lesions of the dorsolateral prefrontal cortex (Brodmann areas 9/46) were impaired in delayed-response tasks with distraction that required active rehearsal, thus involving executive processes. These deficits occurred regardless of lesion side or type of visual information.

In patients with PD, neuropsychological research during the past decade has revealed subtle cognitive deficits in nondemented PD patients apart from the well-known motor symptoms of tremor, rigor, and hypokinesia. Deficits have been shown in "executive" tasks requiring shifting of mental set as in the Wisconsin Card Sorting Test (Canavan et al., 1989a; Heaton, Chelune, Talley, Kay, & Curtiss, 1993; Paolo, Axelrod, Troster, Blackwell, & Koller, 1996) or switching between two concurrent tasks (Downes et al., 1989; Zec et al., 1999). Reduced performance of PD patients in working memory tasks belongs to the most common cognitive deficits. This finding has repeatedly been demonstrated in recent research (e.g., Brown, Soliveri, & Jahanshahi, 1998; Cooper, Sagar, Jordan, Harvey, & Sullivan, 1991; Gabrieli, Singh, Stebbins, & Goetz, 1996; West, Winocur, Ergis, & Saint-Cyr, 1998) and has been attributed to the abovementioned active, organizing processes. Although related to disease severity (Owen et al., 1992), the cognitive deficits in PD can occur already in the early stages of the disease and in the absence of basic intellectual and memory impairments, although they are likely to affect performance in tests on memory and visuospatial performance (for reviews, see Dubois & Pillon, 1997; Levin & Katzen, 1995).

The mild dysexecutive syndrome in patients with PD is frequently attributed to the degeneration of multiple neuronal interconnections between frontal cortex and the basal ganglia (Alexander, DeLong, & Strick, 1986). Degeneration of cells in the pars compacta of the substantia nigra, which is characteristic for the disease, causes a functional deafferentiation of these dopaminergic frontostriatal loops. Recent neuroimaging research showed that for working memory, the interconnections between prefrontal cortex and the rostral parts of the basal ganglia play an important role (D'Esposito et al., 1999; Müller, Wächter, Barthel, Reuter, & von Cramon, 2000; Owen, Doyon, Dagher, Sadikot, & Evans, 1998).

Studies directly comparing patients with PD and FR are rare. Rogers et al. (1998) examined both groups with a task-switch paradigm and found that patients with left-sided but not right-sided frontal damage exhibited markedly increased time costs in switch trials as compared to nonswitch trials. Patients with PD, in contrast, did not show larger time costs but did show increases of error costs. Canavan et al. (1989b) compared performance of several patient groups, for example, FR and PD patients, in several short-term memory tasks. The FR group made more errors in Digit span backward and sequences of three hand gestures, whereas no impairment was found in Digit span forward, Corsi block tapping, and "gesture span", which required the reproduction of gesture sequences of increasing length. Patients with PD performed as well as controls in all of these short-term memory tasks. However, it has to be considered that the PD samples in both studies were in the first stage of the disease in which cognitive impairments are still subtle and probably more difficult to detect than in FR patients.

Given the importance of working memory in everyday life and the high prevalence of working memory deficits in neurological patients, there is a need for measures that are specific for working memory, easy to apply, appropriate for follow-up investigations, and comparable to other widely used neuropsychological tests to allow for the interpretation of individual test profiles. In the following, we will present a short test that we believe meets these criteria. Based on a paper–pencil test devised by Corkin (1968), the Digit Ordering Test (DOT) was introduced by Cooper et al. (1991). In this test, subjects listen to series of seven digits in random order and are asked to immediately recall them in ascending numerical order. The entire test consists of 15 digit sets. Presentation rate is seven digits in 5 s. The test bears a certain resemblance to the Digit span forward and backward subtests of the Wechsler tests (Wechsler, 1987, 1997) in which series of digits have to be repeated in the original or in reversed order. However, it differs from the Digit spans in several aspects: (a) it requires mental reorganization according to an overlearned ordering principle; (b) it is not adaptive with respect to item length and discontinuation after failure; and (c) presentation rate is slightly faster.

Cooper and Sagar (1993), Cooper et al. (1991), Cooper, Sagar, and Sullivan (1993), and Cooper, Sagar, Tidswell, and Jordan (1994) conducted a series of studies with PD patients showing that DOT performance was selectively reduced while Digit spans remained unimpaired. Additionally, Cooper et al. (1992) could demonstrate the sensitivity of DOT performance for detecting changes in working memory capacity following pharmacological treatment in PD. Recently, the findings of Cooper et al. in medicated PD patients have been replicated with the original test version (Gabrieli et al., 1996; Hoppe, Müller, Werheid, Thöne, & von Cramon, 2000).

However, the psychometric evaluation by Hoppe et al. (2000) revealed that the difficulty of the Cooper test was too high, as the majority of them required negative feedback (PD group: 79%; control group: 62%). All test items were of supraspan length and put high demands on both short-term memory and working memory. Frequent negative feedback caused motivational problems in patients performing the test, which were confounded with cognitive deficits. Moreover, item difficulty varied substantially despite of the constant length of all items, which was probably due to unbalanced structural properties (e.g., double digits) throughout the test. Hoppe et al. introduced an experimental modification (DOT-EXP) with items of increasing length in analogy to the Wechsler Digit spans and with balanced structural properties. Experimental variation of the presentation rate revealed that differences between patients with PD and controls were most pronounced at standard presentation rate. Based on these findings, we developed the DOT-A. The test incorporates the features proven most favorable in the DOT-EXP study, i.e., items of balanced structure and increasing difficulty administered at standard presentation rate. It was developed to provide an economic measure for assessment of working memory with maximal comparability to the Wechsler Digit spans.

In the following, we report results from two studies, which represent the first applications of the DOT-A. In the first study, DOT-A and Digit span performance was assessed in PD and FR patients and two groups of healthy subjects pairwise matched to the patients for sex, age, and years of education. The purpose of the study was to investigate whether the findings of Cooper et al. (1991, 1992, 1993, 1994) and Hoppe et al. (2000) in PD patients could be replicated and extended on FR patients. In the second study, a representative sample of healthy subjects was examined in order to obtain validity and reliability estimates of the test.

2. Study 1: Clinical application

2.1. Methods

2.1.1. Subjects

Two clinical groups were examined in this study: 20 patients with PD and 13 patients with FR. We included only PD patients who (a) did not meet diagnostic criteria for dementia accord-

Table 1
Study 1: Demographic data of patients with PD, patients with FR, and control groups

	PD	PD controls	FR	FR controls
Sample size (*n*)	20	20	13	13
Females/males	11/9	11/9	7/6	7/6
Age [years; *M* (S.D.)]	58.2 (7.8)	58.8 (7.3)	39.7 (13.9)	39.5 (13.9)
Years of education [*M* (S.D.)]	13.7 (3.5)	13.2 (2.2)	13.1 (2.5)	14.3 (2.1)

ing to DSM-IV, (b) ranged above the cutoff score of 24 in the Mini Mental Status Examination (Folstein, Folstein, & McHugh, 1975), and (c) were able to participate in a neuropsychological investigation of approximately 1 h. Disease severity, as assessed by an experienced neurologist, ranged between Stages 1 and 4 of the modified Hoehn and Yahr scale (Fahn & Elton, 1987; $M = 2.0$; S.D. = 0.76; Modus: Stage 2). Accordingly, patients showed relatively mild motor symptoms as quantified by the motor scale of the Unified Parkinson's Disease Rating Scale (UPDRS; Martinez-Martin et al., 1994; $M = 25.21$; S.D. = 10.06; Range 9–43). All patients were under medication at the time of testing: Nine patients were exclusively receiving dopamine agonists, nine patients were under combination therapy (L-DOPA and dopamine agonist), and two patients were taking budipine.

The FR group consisted of 13 patients from the Day Care Clinic of Cognitive Neurology at the University of Leipzig who exhibited lesions mainly involving the prefrontal cortex. The patients were free from aphasia, acalculia, and degenerative diseases. Etiologies were traumatic brain injury ($n = 9$), ischemic stroke ($n = 2$), and subarachnoidal hemorrhage caused by aneurysm of the anterior communicating artery ($n = 2$). Lesion sites were confirmed by analysis of individual high-resolution magnetic resonance images.

As the patient groups differed in age [$t(31) = 4.92$, $P < .000$], two separate groups of healthy controls were matched pairwise to the patients for age, sex, and years of education. All subjects were Caucasians and free from neurological disease. Demographic data are presented in Table 1. Controls were recruited through advertisements in the local newspaper and received $6 per hour for their participation. All subjects participated voluntarily and gave written informed consent.

2.1.2. Materials

Digit spans forward and backward were administered according to the assessment guidelines of the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987). The DOT-A consists of six items of increasing length (three to eight digits). Each item comprises two trials, one of them containing a double digit. Subjects are asked to repeat these digits in ascending order immediately after presentation. Presentation rate (one digit per second) and scoring are analogous to the Wechsler spans. As in the Digit spans, the test is discontinued after failure on both trials of any item. Thus, the test differed from the Wechsler spans only with respect to the instruction to order the digits. A test form is provided in the Appendix.

2.1.3. Procedure

Digit spans and DOT-A were administered in the same test session by trained examiners. As pilot testing had revealed that instruction comprehension in patients was optimal when the

DOT-A was administered after the Digit spans, the order of presentation was held constant: Digit span forward, Digit span backward, DOT-A. Administration time for the three tests was 10–15 min.

Test scores were calculated in analogy to the WMS, scoring one point for any correct item. For direct comparison of performance in DOT-A and Digit spans, we additionally determined maximal span length by scoring the number of digits of the last correctly resolved item. If only one trial of the longest item was answered correctly, 0.5 points were subtracted. For example, if only one trial of item length 5 was answered correctly, maximal span length would be 4.5.

2.1.4. Statistical analysis

All statistical analyses were carried out using the Statistical Package for Social Science (SPSS for Windows 6.0). For both performance indices, mixed-factor MANOVAs were conducted with “group” (4) as between-factor and “test” (3) as within-factor. To test the a priori hypothesis of selectively reduced DOT-A performance in the patient groups, t tests of group and test differences were performed. Significance level was set at $P < .05$. For calculation of effect sizes, Cohen’s d was used (Cohen, 1987).

2.2. Results

Mean test scores for Digit spans and DOT-A as well as effect sizes of the group differences are presented in Table 2. The mixed-factor MANOVA yielded significant main effects for Group [$F(3, 62) = 3.30$, $P = .026$] and Test [$F(2, 124) = 22.42$, $P < .001$] as well as a significant Group × Test interaction [$F(6, 124) = 4.44$, $P < .001$]. Pairwise t tests of group differences revealed that patients with PD differed from their control group only in the DOT-A [$t(38) = 4.0$, $P < .001$] but not in the span measures. Within-group comparisons of test differences showed that for the PD group, DOT-A score was significantly lower than Digit span forward score [$t(19) = -5.41$, $P < .001$] but did not differ statistically from Digit span backward. Conversely, for the PD-C group, DOT-A score was significantly higher than Digit span backward [$t(19) = 2.28$, $P = .03$] but did not differ significantly from Digit span forward. For comparison of the FR and FR-C group, the same analyses were performed. Single t testing revealed that FR patients differed from their controls only in the DOT-A [$t(24) = 2.44$, $P = .024$] but not in Digit span backward and Digit span forward. For the FR group, again, DOT-A performance was significantly lower than Digit span forward [$t(12) = -2.71$, $P = .019$] but did not differ from Digit span backward. For the FR-C group, DOT-A and

Table 2
Study 1: Test scores and of patients with PD, patients with FR, and control groups

	PD *M* (S.D.)	PD controls *M* (S.D.)	*d*	FR *M* (S.D.)	FR controls *M* (S.D.)	*d*
Digit span forward	7.9 (1.7)	8.2 (1.6)	0.18	8.1 (1.4)	8.0 (1.8)	0.06
DOT-A	5.5 (1.6)	7.9 (2.2)	1.25	6.6 (2.2)	8.4 (1.4)	0.98
Digit span backward	6.1 (1.5)	7.0 (1.5)	0.6	6.4 (2.0)	6.4 (1.7)	0.00

Group differences are reported as effect sizes (d) in standard deviations.

Digit span backward scores differed statistically [$t(12) = 3.53$, $P = .017$] but not DOT-A and Digit span forward.

In accordance with these findings, effect sizes above one standard deviation were only present in DOT-A group differences. Thus, the test score distributions of DOT-A in patients and respective control groups showed less overlap than test score distribution of the Digit spans. Testing of group differences revealed the same results for maximal spans and test scores. Means and standard deviations of maximal span length for each group are depicted in Figure 1.

To investigate if the pattern of selectively reduced DOT-A performance in patients was also present in single cases, individual data sets were examined. For each individual, "ordering

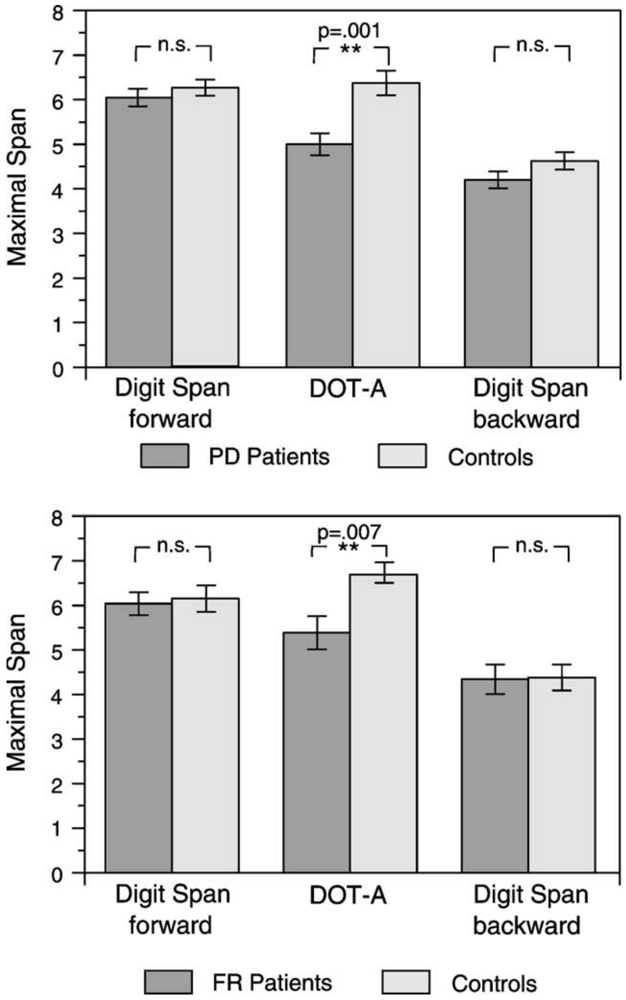

Fig. 1. Means of maximal span length in the DOT-A, Digit spans forward and backward. Upper part: patients with PD and matched controls. Lower part: patients with FR and matched controls. Error bars show the standard error of means.

costs" were determined by subtracting DOT-A score from Digit span forward score. Residuals above zero indicated that DOT-A performance was below Digit span forward performance. Sensitivity and specificity of these ordering costs were further analyzed under the assumption that working memory deficits were present in patients but not in controls. The presence of ordering costs in the two patient groups was taken as an index of sensitivity, i.e., the ability to detect working memory deficits in patients. We found ordering costs in 85% of the PD patients and 62% of the FR patients. On the other hand, the absence of ordering costs in healthy subjects served as an index of specificity. About 40% of the PD-C and 38% of the FR-C group exhibited ordering costs. A nonparametric χ^2 test examining the frequency distribution of ordering costs by group (patients vs. controls) was significant [$\chi^2 = 8.93$]. For a comparison, we also determined "reversal costs" by subtracting Digit span forward score from Digit span backward. Reversal costs were found in 75% of the PD patients, 92% of FR patients, 85% of FR controls, and 75% of PD controls. Thus, lower performance in Digit span backward as compared to Digit span forward were not at all specific for patients. Consequently, the χ^2 did not reach significance [$\chi^2 = 0.1$].

2.3. Discussion

Study 1 revealed selectively reduced DOT-A performance in both patient groups as compared to controls. In patients, DOT-A scores averaged on a similar level as Digit span backward, whereas in controls, DOT-A and Digit span forward scores ranged at the same level. This relative DOT-A performance deficit seems to be particularly sensitive for patients with PD. The difficulties of PD patients in DOT-A have to be differentiated from general cognitive decline in dementia. In a recent review on dementia in PD, Zakzanis and Freedman (1999) pointed out that there were no fundamental differences in the cognitive profiles of demented and nondemented PD patients. The prevalence of dementia increased with disease severity. While generally agreeing with the continuum hypothesis of dementia in PD, we want to point out that the clinical characteristics of the patients included in our sample were highly similar to those of the nondemented patients in the studies reviewed by Zakzanis and Freedman.

It might be argued that a lower DOT-A performance as compared to Digit span forward could be due to instruction complexity. Therefore, we reexamined patients' test protocols and found that all of the patients responded correctly to the first trial of the first item (three digits). Additionally, we searched for errors arising from incorrect numerical ordering, i.e., answers containing a digit that was followed by a lower one. This type of error was very rare, it occurred only once in the PD and once in the FR group. It can be concluded that the concept of numerical order was available for all patients. Consequently, we can rule out that the DOT-A deficit arises simply from difficulties in following the instructions.

DOT-A appears to have a higher sensitivity for detecting working memory deficits in patients with FR and PD compared to Digit span forward and backward. "Ordering costs," i.e., lower performance in DOT-A as compared to Digit span forward, were more frequently found in patients than in healthy controls. Thus, we replicated the findings of Hoppe et al. (2000) of a single dissociation of reduced verbal working memory and intact short-term memory in PD patients. Additionally, the same pattern of performance was found in patients with FR. The similarity of findings supports the hypothesis that functional deafferentiation of the frontal cortex is the cause of the cognitive deficits in PD. Finally, the results of Study 1 demonstrate that DOT-A is well suited for the examination of patient groups with respect to cognitive demands and test economy.

3. Study 2: Concurrent validity and reliability

For further examination of validity and reliability of DOT-A, a second study with a larger sample of healthy subjects was conducted.

3.1. Methods

3.1.1. Subjects

In order to obtain maximal representativity for the population, subjects were carefully selected according to age, sex, education, and handedness. First, an age range between 20 and 69 was determined for the entire sample, which was subdivided into five decades. For each decade, 10 persons (five females) were assessed. The proportion of left handers and the sum of years of formal education equaled the proportion in the respective age cohort of the German population as reported by the German Ministry for Education and Research (BMB + F). All subjects were Caucasians and free from neurological disease. They were recruited by newspaper advertisements and received $6 per hour for their participation. Control subjects from Study 1 were included in the sample. Demographic characteristics are shown in Table 3.

3.1.2. Materials

Digit spans and DOT-A were employed as in Study 1. A parallel DOT version, the DOT Form B (DOT-B), was constructed for investigation of parallel test reliability. Item structure (e.g., prevalence of double digits) was conserved by applying a simple additive shift (+1 or −1) on the DOT-A items. A test form is given in the Appendix. The subtest “Working Memory” as implemented in the computerized “Test Battery for the Assessment of Attention” (TAP; Zimmermann & Fimm, 1993) was employed for measurement of concurrent validity. In this computerized two-back task, single one-figure digits between 1 and 9 are presented on the screen for 1.5 s with an interstimulus interval of 1.5 s. Subjects are instructed to press a key if a given number equals the number presented before its antecedent. To accomplish this, they have

Table 3
Study 2: Demographic data of normative sample

	Age groups					
	20–29	30–39	40–49	50–59	60–69	Total sample
Sample size (*n*)	10	10	10	10	10	50
Male/female	5/5	5/5	5/5	5/5	5/5	25/25
Age [years; *M* (S.D.)]	22.3 (2.6)	34.2 (2.9)	45.7 (3.1)	55.9 (2.4)	64.0 (3.3)	44.4 (15.3)
Formal education [years; *M* (S.D.)]	10.8 (1.0)	11.5 (1.4)	10.2 (1.5)	9.8 (1.5)	9.4 (1.6)	10.3 (1.5)
Total years of education [*M* (S.D.)]	13.2 (0.6)	15.4 (1.7)	13.4 (2.6)	13.0 (2.5)	13.4 (2.3)	13.7 (2.2)

to monitor each number and simultaneously keep in mind the two numbers having appeared before. Reaction times on correct trials, number of misses, and false positive responses are recorded.

3.1.3. Procedure

DOT-A and DOT-B were administered in separate test sessions with a retest interval of 8–10 days. A crossover design was applied by assigning half of the sample (five subjects per subgroup) to completion of DOT-A in the first session and DOT-B in the second session. The other half received DOT-B first. The two-back task was always completed in the second test session.

3.1.4. Statistical analysis

Differences between mean scores of DOT-A and Digit spans as well as differences in DOT-A performance between age groups were examined by a one-way ANOVA. To test the a priori hypothesis of equal Digit span forward and DOT-A performance in healthy subjects (cf. Study 1; Cooper et al., 1991; Hoppe et al., 2000), *t* tests of differences between mean scores in these tests were performed. Significance level was set at $P < .05$. For all correlational analyses, parametric correlation coefficients (Pearson's *r*) were computed. Wilk's λ (Wilks, 1946) was computed for examining if the tests were parallel. The Spearman–Brown formula was employed for calculation of split-half reliability. For calculation of parallel test reliability (r_{par}), we used the Cureton formula (1971) (see Appendix) in which correlation between two tests ($r_{(\text{DOT-A,DOT-B})}$) is corrected for the attenuation caused by imperfect reliabilities of each single test, in this case, split-half reliabilities $r_{\text{tt(DOT-A)}}$ and $r_{\text{tt(DOT-B)}}$.

3.2. Results

3.2.1. Test scores

Mean test scores of DOT-A, DOT-B, both Digit spans, and the two-back task for each age group and the total sample are presented in Table 4. A one-way ANOVA revealed a significant main effect for test [$F(2, 49) = 15.16$, $P < .001$]. Pairwise *t* testing revealed that the difference between DOT-A and Digit span forward did not reach significance [$t(49) = -1.97$, $P = .06$], but significant differences were found between DOT-A and Digit span backward [$t(49) = -3.05$, $P = .004$].

3.2.2. Parallelism of DOT-A and DOT-B

For simultaneous examination of means and variances, Wilk's λ was calculated for DOT-A and DOT-B. The resulting χ^2 value was not significant [$\chi^2 = 0.836$]. Thus, means, variances, and covariance of DOT-A and DOT-B did not differ significantly, indicating that the hypothesis of parallel tests could be accepted for the entire normative sample.

3.2.3. Reliability

Firstly, split-half reliability was computed by correlating the test halves. As the second trial of any item contained a double digit, lower performance in every second trial was likely to attenuate reliability coefficients artificially. Therefore, instead of the odd–even method of item

Table 4
Study 2: Test scores of normative sample

	Age groups					
	20–29 *M* (S.D.)	30–39 *M* (S.D.)	40–49 *M* (S.D.)	50–59 *M* (S.D.)	60–69 *M* (S.D.)	Total sample *M* (S.D.)
Digit span forward	9.6 (1.6)	8.3 (1.2)	8.8 (1.8)	7.9 (1.2)	8.2 (1.8)	8.6 (1.6)
DOT-A	8.3 (1.7)	8.6 (1.8)	7.9 (2.3)	7.6 (2.3)	7.5 (1.9)	8.0 (2.0)
DOT-B	9.6 (1.3)	8.5 (1.8)	8.1 (2.2)	7.0 (1.7)	7.8 (1.6)	8.2 (1.9)
Digit span backward	8.4 (1.6)	6.9 (2.0)	6.3 (1.8)	6.3 (1.1)	7.2 (1.5)	7.0 (1.7)
Two-back task (RT)	656.8 (256.8)	614.1 (187.8)	595.0 (167.2)	686.5 (172.7)	602.9 (128.3)	631.0 (183.0)
Two-back task (omissions)	2.6 (2.6)	2.0 (1.7)	1.7 (1.9)	1.1 (1.1)	2.3 (2.8)	1.9 (2.1)
Two-back task (false positives)	2.4 (2.1)	1.9 (1.8)	4.8 (2.5)	6.7 (7.6)	4.2 (4.8)	4.0 (4.5)

RT: median of reaction times for correct responses.

assignment, we used a crossover item assignment, i.e., the first test half included the first trial of Item 1, the second trial of Item 2, and so forth.

Spearman–Brown coefficients were $r_{\mathrm{tt(DOT\text{-}A)}} = .75$ and $r_{\mathrm{tt(DOT\text{-}B)}} = .79$. In order to examine the effect of test prolongation, split-half reliability was also determined for the mean of DOT-A and DOT-B scores in both test halves. The resulting Spearman–Brown coefficient was $r_{\mathrm{tt(DOTA/B)}} = .86$. As expected, Spearman–Brown coefficients were lower when the tests were subdivided according to the odd–even method ($r_{\mathrm{t1t2(DOT\text{-}A)}} = .70$ and $r_{\mathrm{t1t2(DOT\text{-}B)}} = .66$). As a second reliability estimate, parallel test reliability for DOT-A and DOT-B was determined by applying the Cureton (1971) formula (see Appendix). Parallel test reliability was $r_{\mathrm{par}} = .68$.

3.2.4. Concurrent validity

Correlation coefficients were calculated for DOT-A, DOT-B, and three parameters of the two-back task, namely median reaction times, omissions, and false positives. DOT-A and DOT-B scores were negatively correlated with omission errors in the two-back task [$r_{\mathrm{(DOT\text{-}A,\,omissions)}} = -.33$, $P = .019$; $r_{\mathrm{(DOT\text{-}B,\,omissions)}} = -.29$, $P = .040$] as well as with false positive responses of the two-back task [$r_{\mathrm{(DOT\text{-}A,\,false\,positives)}} = -.41$, $P = .003$; $r_{\mathrm{(DOT\text{-}B,\,false\,positives)}} = -.28$, $P = .05$]. However, there was no significant correlation with performance speed [$r_{\mathrm{(DOT\text{-}A,\,median\,reaction\,times)}} = -.04$, $P = .781$; $r_{\mathrm{(DOT\text{-}B,\,median\,reaction\,time)}} = -.14$, $P = .318$].

3.2.5. Age effects

Correlation with age was significant for both Digit spans [$r_{\mathrm{(Digit\,span\,forward,\,age)}} = -.30$, $P = .033$; $r_{\mathrm{(Digit\,span\,backward,\,age)}} = -.29$, $P = .041$]. A significant age decline was also found for DOT-B [$r_{\mathrm{(DOT\text{-}B,\,age)}} = -.44$, $P = .001$] but not for DOT-A [$r_{\mathrm{(DOT\text{-}A,\,age)}} = -.20$, $P = .16$]. Examination of the descriptive data suggests that this was mainly due to relatively low performance of the youngest age group in DOT-A, with half of the group obtaining a maximal span length of only five or six items. To explore if this finding was due to limited sample sizes of $n = 10$ per age group, the means of DOT-A and DOT-B scores were calculated for each group. Correlation with age was significant for this hypothetical DOT mean score [$r = -.30$, $P = .010$].

Table 5
Study 2: Distribution of DOT-A scores in healthy subjects ($N = 50$)

Test score	Percentile	
	DOT-A	DOT-A/B
12	98	98
11	93	92
10	81	81
9	66	65
8	55	47
7	37	28
6	16	13
5	5	5
≤4	1	1

3.2.6. *Standardization*

As sample sizes of the subgroups appeared not to be large enough for standardization and mean scores in the five age groups did not differ significantly, percentile scores were calculated for the entire sample (see Table 5).

3.3. *Discussion*

In Study 2, previous findings on the relationship between Digit span forward and DOT-A performance in healthy subjects were replicated in a larger sample. Examination of parallel test equivalence for the entire sample revealed that DOT-A and DOT-B are parallel tests. Split-half reliability for both tests ranged at .75 for DOT-A and .79 for DOT-B. Thus, the coefficients ranged slightly below the split-half reliability coefficients of the Digit spans according to the manual of the WMS-R (Wechsler, 1987), which average between .80 to .85 for the different age groups. Split-half reliability of the DOT scores could be enhanced to this level by test prolongation, when the mean score of DOT-A and DOT-B test halves was considered. Parallel test reliability was ranging slightly lower than split-half reliability at .68. This is consistent with the view that parallel test reliability is a more conservative reliability estimate. The difference between split-half reliability and parallel test reliability can be attributed to the testing conditions, which are stable for the former and variable for the latter. In summary, reliability can be considered as sufficient.

Examination of age effects showed that DOT-A and DOT-B test scores were not consistently correlated with age (cf. Dobbs & Rule, 1989). We attribute this finding to the small sample sizes of the age groups. Consequently, when doubling sample sizes by correlating the mean of both tests with age, the relationship achieved significance. This fits well with the finding that tests were parallel according to the Cureton formula, which incorporates the results of total sample. We conclude that the moderate size of the normative sample does not allow for further subdivision and therefore computed percentile scores for the entire sample.

Both parallel test forms could be externally validated as a measure of working memory by moderate but significant negative correlations with the error scores of a two-back working memory task. Interestingly, correctness but not speed of answering in the computerized two-back task was related to the DOT scores. Consistent with this finding, median reaction times, and error scores of the two-back task were not correlated. Thus, there is neither a speed–accuracy tradeoff nor a coupling of high speed and correct performance in healthy subjects' two-back task results. This finding has two implications: Firstly, we can rule out that the correlations between two-back task error scores and DOT-A are purely relying on a response bias. Secondly, in a more general view, examiners should focus on error scores instead of reaction time indices for the interpretation of patients' performance in a two-back task. We can add that this finding corresponds to our clinical experience with this type of task.

The promising results of the psychometric analysis in the present study are limited by relatively small sample sizes. There is no doubt that despite careful sample selection, it can only provide preliminary data on the psychometric test features of DOT-A. Further studies should be conducted to examine larger samples. Nevertheless, from a clinical point of view, it is encouraging that consistent results can already be found in small samples.

4. General discussion

Results from two studies were presented, which examined the DOT-A as a measure of verbal working memory in analogy to the Wechsler Digit spans. Our results are in line with previous studies that revealed a dissociation between two working memory components: storage or short-term memory and executive, information-manipulating processes. Clinical application of the test confirmed earlier findings (Hoppe et al., 2000) that DOT-A particularly involves the manipulation component of working memory. The analysis of means and effect sizes revealed that substantial differences between patients and healthy controls were only found in DOT-A but not in the Digit span measures. This finding is remarkable, as the test is highly analogous to the Digit spans and differs only with respect to the instruction.

Concerning Digit span backward, our findings are in contrast with several previous studies. In neither of the patient samples of Study 1, we did find a significant reduction of Digit span backward performance as compared to controls. However, Cooper et al. (1991) found reduced performance in PD patients in both Digit span backward and the original DOT version. Canavan et al. (1989b) found Digit span backward to be impaired in FR patients. In many clinical studies and in diagnostic practice, Digit span backward has been considered as a working memory test (e.g., Zakzanis, Leach, & Kaplan, 1999). Our results contradict this view. However, they are in line with the findings of Rosen and Engle (1997), who compared forward and backward letter spans and found that they did not differ in complexity, in format of mental representation, or in predicting intellectual ability in a student population. In a factor analysis by Engle et al. (1999), backward word span loaded together with forward span is highest on a factor interpreted as representing short-term memory and not on the working memory factor. The common finding that backward spans are lower than forward spans ("reversal costs") could probably be explained by higher demands on the storage component, as the time elapsed between presentation and reproduction of the first digits is consistently longer than in forward span. On the basis of our findings, we conclude that the common reduction of Digit span backward performance compared to Digit span forward is primarily due to higher demands on short-term memory.

The question arises, what turns the DOT-A to a working memory test? Subjects have to apply an overlearned ordering principle on a given memory set, which has to be held in memory until ordering has been finished. Taken together with previous findings (Hoppe et al., 2000), which revealed that higher or lower presentation rate reduced DOT performance, we conclude that the test involves features of a dual task. PD and FR patients' difficulties arise neither from general cognitive slowing nor from impaired short-term memory or from lacking availability of the ordering routine. Rather, they are impaired in parallelizing storage and ordering processes within a constant time interval defined by presentation rate. This deficit in parallel processing could be due to a lack of internally guided use of strategic memory routines.

What are the advantages of DOT-A in comparison with other working memory tasks, e.g., the Alphabetical Word Span Test (Belleville, Rouleau, & Caza, 1998), the Day of the Week Ordering Test (Kempler, Almor, Tyler, Andersen, & Macdonald, 1998), Letter–Number Sequencing (WAIS-III; Wechsler, 1997), or the Paced Auditory Serial Attention Test (PASAT; Gronwall & Wrightson, 1981)? Firstly, the analogy to the Wechsler Digit spans allows for the investigation of short-term memory and working memory with the same materials. As we could show in our analysis of "ordering costs," direct comparison of Digit spans and DOT-A allows to dissociate between individual working memory and short-term memory capacity in clinical practice. Secondly, DOT-A appears to be a relatively pure measure of working memory. Compared with other ordering measures, like Letter–Number Sequencing, it bears the advantage of maximal comparability with the Wechsler spans, as the same numerical material is used. Finally, DOT-A seems to be particularly sensitive for patient populations. It would be interesting to compare individuals with high or low spans in other working memory tasks to investigate the question if individuals with low spans perform in DOT-A on the same level as patients in our study did. However, the most promising application of the DOT-A would be in the clinical context, as alternative working memory tasks, e.g., the PASAT, are sometimes too difficult, time-consuming, and exhausting for patients with reduced endurance.

The importance of working memory for patients' everyday functioning, such as verbal communication or activities of daily living, is immense. Therefore, the wealth of recent experimental evidence on working memory should be embodied in clinical practice by deriving reliable and valid instruments for neuropsychological diagnosis and evaluation. We hope that we can contribute to this purpose by the present, tentative analyses on the DOT.

Acknowledgment

This research was supported by the German Bundesministerium für Bildung und Forschung (BMB + F), Interdisciplinary Centre for Clinical Research at the University of Leipzig (01K59504, project C09). We would like to thank Mike Reuter, MD, for referring his patients and providing clinical ratings. We are also grateful to Laura Busse, Grit Hein, and Gundula Jänicke who participated in collection and analysis of the data and to Evelyn Ferstl, PhD and Kathrin Walther for useful methodological advice.

Appendix

Version A, Adaptive Digit Ordering Test (DOT-A)

Patient ID:
Age: m f
Date:
Examiner:

Instruction: In the following test, you will be presented several digits in random order. Please recall these digits in ascending order immediately after presentation. If digits occur twice, recall them twice as well.

Example: 5 - 2 - 8 - 2; *Solution:* 2 - 2 - 5 - 8

Assessment: Read one digit per second with equal stressing and prosody, drop the pitch of your voice on the last digit of each trial. Discontinue if both items of any item are wrong. If possible, assess Digit span forward and backward in the same session. Protocolling answers is facultative.

Scoring: (a) *Test score:* 1 point for each correct answer; (b) *Maximal span:* length of the last correctly reproduced item. Subtract 0.5 points if only one item on this level has been answered correctly.

Digit Ordering-A

Item	Span length	Trial 1	1/0	Trial 2	1/0	Score (0–2)
1.	*3*	3 7 2 (2 3 7)		6 1 6 (1 6 6)		
2.	*4*	8 4 7 3 (3 4 7 8)		7 2 7 6 (2 6 7 7)		
3.	*5*	2 6 5 3 1 (1 2 3 5 6)		5 1 8 5 2 (1 2 5 5 8)		
4.	*6*	4 8 3 9 2 7 (2 3 4 7 8 9)		3 0 3 7 2 5 (0 2 3 3 5 7)		
5.	*7*	7 4 3 1 8 2 6 (1 2 3 4 6 7 8)		9 6 5 2 8 6 3 (2 3 5 6 6 8 9)		
6.	*8*	3 7 6 1 8 2 5 0 (0 1 2 3 5 6 7 8)		2 6 3 2 1 8 7 5 (1 2 2 3 5 6 7 8)		
Notes:				Test score (max. 12): Percentile of test score: Maximal span:		

Version B, Adaptive Digit Ordering Test (DOT-B)

Patient ID:
Age: m f
Date:
Examiner:

Instruction: In the following test, you will be presented several digits in random order. Please recall these digits in ascending order immediately after presentation. If digits occur twice, recall them twice as well.

Example: 5 - 2 - 8 - 2; *Solution:* 2 - 2 - 5 - 8

Assessment: Read one digit per second with equal stressing and prosody, drop the pitch of your voice on the last digit of each trial. Discontinue if both items of any item are wrong. If possible, assess Digit span forward and backward in the same session. Protocolling answers is facultative.

Scoring: (a) *Test score:* 1 point for each correct answer; (b) *Maximal span:* length of the last correctly reproduced item. Subtract 0.5 points if only one item on this level has been answered correctly.

Digit Ordering-B

Item	Span length	Trial 1	1/0	Trial 2	1/0	Score (0–2)
1.	*3*	4 8 3 (3 4 8)		7 2 7 (2 7 7)		
2.	*4*	9 5 8 4 (4 5 8 9)		6 1 6 5 (1 5 6 6)		
3.	*5*	3 7 6 4 2 (2 3 4 6 7)		6 2 9 6 3 (2 3 6 6 9)		
4.	*6*	3 7 2 8 1 6 (1 2 3 6 7 8)		4 1 4 8 3 6 (1 3 4 4 6 8)		
5.	*7*	6 3 2 0 7 1 5 (0 1 2 3 5 6 7)		8 5 4 1 7 5 2 (1 2 4 5 5 7 8)		
6.	*8*	4 8 7 2 9 3 6 1 (1 2 3 4 6 7 8 9)		1 5 2 1 0 7 6 4 (0 1 1 2 4 5 6 7)		
Notes:				Test score (max. 12): Percentile of test score: Maximal span:		

The Cureton (1971) formula for calculation of parallel test reliability:

$$r_{\text{par}} = \frac{r(\text{Test A},\ \text{Test B})}{\sqrt{r(\text{Test A},\ \text{reliability of Test A}) \times r(\text{Test B},\ \text{riability of Test B})}}$$

References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organisation of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neurosciences*, *9*, 357–381.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Belleville, S., Rouleau, N., & Caza, N. (1998). Effect of normal aging on the manipulation of information in working memory. *Memory and Cognition*, *26*, 572–583.

Brown, R. G., Soliveri, P., & Jahanshahi, M. (1998). Executive processes in Parkinson's disease—random number generation and response suppression. *Neuropsychologia*, *36*, 1355–1362.

Canavan, A. G. M., Passingham, R. E., Marsden, C. D., Quinn, N., Wyke, M., & Polkey, C. E. (1989a). The performance on learning tasks of patients in the early stages of Parkinson's disease. *Neuropsychologia*, *27*, 141–156.

Canavan, A. G. M., Passingham, R. E., Marsden, C. D., Quinn, N., Wyke, M., & Polkey, C. E. (1989b). Sequencing ability in Parkinsonians, patients with frontal lobe lesions and patients who have undergone unilateral temporal lobectomies. *Neuropsychologia*, *27*, 787–798.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Hillsdale: Erlbaum.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, *123*, 354–373.

Cooper, J. A., & Sagar, H. J. (1993). Encoding deficits in untreated Parkinson's disease. *Cortex*, *29*, 251–265.

Cooper, J. A., Sagar, H. J., Jordan, N., Harvey, N. S., & Sullivan, E. V. (1991). Cognitive impairment in early, untreated Parkinson's disease and its relationship to motor disability. *Brain*, *114*, 2095–2122.

Cooper, J. A., Sagar, H. J., Doherty, S. M., Jordan, N., Tidswell, P., & Sullivan, E. V. (1992). Different effects of dopaminergic and anticholinergic therapies on cognitive and motor function in Parkinson's disease. *Brain*, *115*, 1701–1725.

Cooper, J. A., Sagar, H. J., & Sullivan, E. V. (1993). Short-term memory and temporal ordering in early Parkinson's disease: Effects of disease chronicity and medication. *Neuropsychologia*, *31*, 933–949.

Cooper, J. A., Sagar, H. J., Tidswell, P., & Jordan, N. (1994). Slowed central processing in simple and go/no-go reaction time tasks in Parkinson's disease. *Brain*, *117*, 517–529.

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, *6*, 255–265.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cureton, E. E. (1971). The stability coefficient. *Educational and Psychological Measurement*, *31*, 45–55.

D'Esposito, M., & Postle, B. R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia*, *37*, 1303–1315.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fMRI study. *Brain and Cognition*, *41*, 66–86.

Dobbs, A. R., & Rule, B. G. (1989). Adults' age differences in working memory. *Psychology and Aging*, *4*, 400–503.

Downes, J. J., Roberts, A. C., Sahakian, B. J., Evenden, J. L., Morris, R. G., & Robbins, T. W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: Evidence for a specific attentional dysfunction. *Neuropsychologia*, *27*, 1329–1343.

Dubois, B., & Pillon, B. (1997). Cognitive deficits in Parkinson's disease. *Journal of Neurology*, *244*, 2–8.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, *84*, 253–262.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Fahn, S., & Elton, R. (1987). Unified Parkinson's Disease Rating Scale. In S. Fahn, C. D. Marsden, D. B. Goldstein, & M. Goldstein (Eds.), *Recent developments in Parkinson's disease* (Vol. 2, pp. 293–304). Florham Park, NJ: Macmillan.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental-state"—a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Gabrieli, J. D. E., Singh, J., Stebbins, G. T., & Goetz, C. G. (1996). Reduced working memory span in Parkinson's disease: Evidence for the role of a frontostriatal system in working and strategic memory. *Neuropsychology*, *10*, 322–332.

Gronwall, D., & Wrightson, P. (1981). Memory and information processing capacity after closed head injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, *44*, 889–895.

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

Hebb, D. O. (1939). Intelligence in man after large removals of cerebral tissue: Report of four left frontal lobe cases. *Journal of General Psychology*, *21*, 73–87.

Hoppe, C., Müller, U., Werheid, K., Thöne, A., & von Cramon, D. Y. (2000). The Digit Ordering Test: Clinical, psychometric, and experimental evaluation of a verbal working memory test. *The Clinical Neuropsychologist*, *14*, 38–55.

Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. C. (1998). Sentence comprehension deficits in Alzheimer's disease: A comparison of off-line vs. on-line sentence processing. *Brain and Language*, *64*, 297–316.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.

Levin, B. E., & Katzen, H. L. (1995). Early cognitive changes and nondementing behavioural abnormalities in Parkinson's disease. In W. J. Weiner & A. E. Lang (Eds.), *Behavioral neurology of movement disorders. Advances in neurology* (Vol. 65, pp. 85–95). New York: Raven Press.

Martinez-Martin, P., Gil-Nagel, A., Gracia, L. M., Gomez, J. B., Martinez-Sarries, J., Bermejo, F., & The Co-operative Multicentric Group. (1994). Unified Parkinson's Disease Rating Scale characteristics and structure. *Movement Disorders*, *9*, 76–83.

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory. In A. Miyake & R. Shah (Eds.), *Models of working memory* (pp. 442–481). Cambridge: Cambridge University Press.

Müller, U., Wächter, T., Barthel, H., Reuter, M., & von Cramon, D. Y. (2000). Striatal [^{123}I]β-CIT SPECT and prefrontal memory functions in Parkinson's disease. *Journal of Neural Transmission*, *107*, 303–319.

Owen, A. M., James, M., Leigh, P. N., Summers, B. A., Marsden, C. D., Quinn, N. P., Lange, K. W., & Robbins, T. W. (1992). Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain*, *115*, 1727–1751.

Owen, A. M., Doyon, J., Dagher, A., Sadikot, A., & Evans, C. (1998). Abnormal basal ganglia outflow in Parkinson's disease identified with PET. Implications for higher cortical functions. *Brain*, *121*, 949–965.

Owen, A. M., Herrod, N. J., Menon, D. K., Clark, J. C., Downey, S. P., Carpenter, T. A., Minhas, P. S., Turkheimer, F. E., Williams, E. J., Robbins, T. W., Sahakian, B. J., Petrides, M., & Pickard, J. D. (1999). Redefining the functional organization of working memory processes within human lateral prefrontal cortex. *European Journal of Neuroscience*, *11*, 567–574.

Paolo, A. M., Axelrod, B. N., Troster, A. L., Blackwell, K. T., & Koller, W. C. (1996). Utility of a Wisconsin Card Sorting Test short form in persons with Alzheimer's and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *17*, 892–897.

Parkin, A. J. (1998). The central executive does not exist. *Journal of the International Neuropsychological Society*, *4*, 518–522.

Petrides, M. (1994). Frontal lobes and working memory: Evidence from investigations of the effects of cortical excisions in nonhuman primates. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 59–82). Amsterdam: Elsevier.

Postle, B. R., Berger, J. S., & D'Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proceedings of the National Academy of Sciences of the United States of America*, *26*, 12959–12964.

Rogers, R. D., Sahakian, B. J., Hodges, J. R., Polkey, C. E., Kennard, C., & Robbins, T. W. (1998). Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease. *Brain*, *121*, 815–842.

Rosen, V. M., & Engle, R. W. (1997). Forward and backward serial recall. *Intelligence*, *25*, 377.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*, 5–42.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*, 1657–1661.

Wechsler, D. (1987). *WMS-R Wechsler Memory Scale-Revised. Manual*. San Antonio: The Psychological Corporation.

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio: The Psychological Corporation.

West, R., Winocur, G., Ergis, A.-M., & Saint-Cyr, J. (1998). The contribution of impaired working memory monitoring to performance of the Self-Ordered Pointing Task in normal aging and Parkinson's disease. *Neuropsychology*, *12*, 546–554.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. *The American Mathematics and Statistics*, *17*, 257.

Zakzanis, K. K., & Freedman, M. (1999). A neuropsychological comparison of demented and nondemented patients with Parkinson's disease. *Applied Neuropsychology*, *6*, 129–146.

Zakzanis, K. K., Leach, L., & Kaplan, E. (1999). *Neuropsychological differential diagnosis*. Lisse: Swets & Zeitlinger.

Zec, R. F., Landreth, E. S., Fritz, S., Grames, E., Hasara, A., Fraizer, W., Belman, J., Wainman, S., McCool, M., O'Connell, C., Harris, R., Robbs, R., Elble, R., & Manyam, B. (1999). A comparison of phonemic, semantic, and alternating word fluency in Parkinson's disease. *Archives of Clinical Neuropsychology*, *14,* 255–264.

Zimmermann, P., & Fimm, B. (1993). *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* (Test battery for the assessment of attention). Würselen, Germany: Psychologische Testsysteme.