Journal of Clinical and Experimental Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713657736

Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and clinical neurology patients

Matthias Henry [a]; Thomas Merten [b]; Simone Andrea Wolf [c]; Sandy Harth [c]
[a] Department of Neurology, Humboldt Klinikum, Berlin, Germany [b] Department of Neurology, Klinikum im Friedrichshain, Berlin, Germany [c] Department of Neurology, Otto von Guericke University, Magdeburg, Germany

First Published on: 08 April 2009

To cite this Article Henry, Matthias, Merten, Thomas, Wolf, Simone Andrea and Harth, Sandy(2009)'Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and clinical neurology patients',Journal of Clinical and Experimental Neuropsychology,

To link to this Article: DOI: 10.1080/13803390902791653

URL: http://dx.doi.org/10.1080/13803390902791653

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2009, i*First*, 1–10

Psychology Press
Taylor & Francis Group

Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and clinical neurology patients

Matthias Henry,[1] Thomas Merten,[2] Simone Andrea Wolf,[3] and Sandy Harth[3]

[1]Department of Neurology, Humboldt Klinikum, Berlin, Germany
[2]Department of Neurology, Klinikum im Friedrichshain, Berlin, Germany
[3]Department of Neurology, Otto von Guericke University, Magdeburg, Germany

The study aimed to provide independent data on the specificity of the Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2008), a new test that combines conventional decision making based on cutoffs with profile analyses in order to identify invalid test performance and to reduce false positive classifications. The results of 65 bona fide neurological patients (with 21 of them meeting *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition,* DSM-IV, core criteria for dementia) were compared to 50 healthy volunteers. One patient was wrongly classified as malingering, resulting in a specificity of 98.5% for neurological patients and 100% for controls. A total of 13 patients with dementia (62%), 6 patients without dementia (14%), and 1 healthy participant exhibited a dementia profile in the NV-MSVT. While these results confirm the high specificity of the NV-MSVT for the classification *insufficient effort*, its sensitivity has to be verified by independent research data.

Keywords: Malingering; Symptom validity testing; Neuropsychological assessment; Negative response bias; Memory.

INTRODUCTION

Symptom validity assessment may be considered a success story of forensic neuropsychology, both in terms of the development of new assessment techniques (Boone, 2007; Larrabee, 2007) and in terms of general acceptance that neuropsychological test results have to be adequately checked for their validity (Bush et al., 2005). The results of independent neuropsychological evaluations may have far-reaching consequences for the parties involved. It may be said that any false decision made by the forensic expert violates the rights of one party. Consequently, classification accuracy is a major concern for all instruments and approaches used for determining negative response bias.

Estimates of sensitivity and specificity of symptom validity tests depend to a large degree upon the specific method employed. Apart from experimental studies with instructed simulators and data collected from forensic patients, it is important to see how bona fide patients with neurological or psychiatric conditions and healthy full-effort controls perform. Ideally, both healthy participants and bona fide patients should pass symptom validity tests (SVTs) provided they put forward full effort. However, true symptomatology, both cognitive and psychiatric, may interfere significantly with SVT performance (cf. Gorissen, Sanz, & Schmand, 2005; Merten, Bossink, & Schmand, 2007). While the presence of authentic psychopathology or cognitive impairment would usually not explain below-chance performance in forced-choice SVTs, the same is not true for performance below empirically established cutoffs. In the criteria for diagnosing malingered neurocognitive symptoms (Slick, Sherman, & Iverson, 1999), this has been the basis for attributing the highest degree of diagnostic confidence to below-chance response patterns. Thus, when using empirical cutoffs for decision making, the problem of false positive identification is imminent (i.e., truly impaired patients may be wrongly classified

Address correspondence to Matthias Henry, Vivantes Humboldt-Klinikum, Neurology, Am Nordgraben 2, D-13509 Berlin, Germany (E-mail: matthiashenry@web.de).

http://www.psypress.com/jcen
DOI: 10.1080/13803390902791653

as malingering). This problem has more recently been discussed on a broader scale, and studies have yielded mixed results, with some findings showing SVTs to be resistant against authentic cognitive impairment and others showing that failure on SVTs may occur due to neurocognitive symptoms (e.g., Batt, Shores, & Chekaluk, 2008; Graue et al., 2007; Heinly, Greve, Bianchini, Love, & Brennan, 2005; Merten et al., 2007; Turner, Horner, Edmiston, & Bachman, 2008).

A hierarchical approach is one method of tackling this problem. This involves analyzing whether the SVT performance of a given patient is above or below a preestablished cutoff point (the conventional pass/fail approach) and then, for patients who fail, investigating the profile over several subtests. With a sufficiently large database, patients with authentic neurocognitive conditions such as dementia, aphasia, or confusional state may be distinguished from malingerers on the basis of their profiles, although both groups may fail the cutoffs for suboptimal performance in the first place. This approach has been especially promoted by Green (2003, 2004, 2008), who developed three SVTs: the Word Memory Test (WMT), the Medical Symptom Validity Test (MSVT), and the Nonverbal Medical Symptom Validity Test (NV-MSVT). In fact, the latter of the three tests has been explicitly developed on the basis of this hierarchical approach.

The test manual gives plenty of information about results obtained with the NV-MSVT so far, which has been gathered by the author and independent investigators around the world. Data are available on different groups of adults and children, honestly performing healthy adult volunteers, and experimental simulators. Also, data from patients with dementia are reported. None of these failed the NV-MSVT classification rules described in the Method section of this article. In a group of patients with moderate or severe traumatic brain injury (TBI), the NV-MSVT failure rate was zero while 30% of claimants with mild TBI failed the test. Investigating diagnostic accuracy of the classification rules, Green (2008) reported a sensitivity estimate for *poor effort* of 72.5%, established with a group of simulators.

In a recent simulator study, Weinborn, Woods, and Fox (2008) also found that both the MSVT and the NV-MSVT were powerful in identifying experimental malingerers. However, currently there exist only limited data on the reliability and validity of the profiles obtained for these tests (Howe & Loring, 2009).

The current study aimed to provide independent data on the specificity of the NV-MSVT and its stepwise approach to assessing suboptimal test effort. Therefore, the classification accuracy in a group of clinical neurology patients and healthy adults was examined to check the empirical cutoffs proposed by Green (2008). For any symptom validity measure, data on clinical groups and normal controls are an important feature, which may substantially facilitate the interpretation of test data obtained in forensic contexts. Furthermore, the study aimed to examine whether the profile analysis proposed by Green can be used to distinguish between genuine cases of dementia and patients without dementia. Also, the association between NV-MSVT scores and neuropsychological test performance was analyzed in order to examine how far the ability to pass the SVT depended on neurocognitive factors.

METHOD

Research participants

Two groups of native German-speaking adults were investigated. The first group consisted of 65 neurological bona fide patients who underwent clinical neuropsychological assessment at the University of Magdeburg Medical Center, Department of Neurology, from June 2006 to August 2008. Patients were excluded if severe perceptional or language deficits were present or if attentional resources were severely limited; thus, patients with confusional states or severe dementia were not included. The clinical diagnoses of the participants could be summarized as follows: 31% suffered from cerebrovascular diseases, 46% degenerative, 5% neoplastic, 5% normal pressure hydrocephalus, 3% primary epilepsy, 3% traumatic brain injury, 3% inflammatory diseases, and 5% diverse.

The patients were divided into two subgroups according to their cognitive states: patients with dementia and patients who did not meet criteria for dementia. For this decision, DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition*; American Psychiatric Association, 1995) core criteria for dementia were applied: (a) development of multiple cognitive deficits manifested by both memory and other cognitive disorders; (b) significant impairment in social or occupational functioning; significant decline from previous level of functioning; (c) deficits not occurring exclusively during the course of a delirium; (d) impairment not better accounted for by another Axis I disorder. A total of 21 of the patients (32%) met these criteria.

The subgroup of patients with dementia was composed of cases of Alzheimer's disease (n = 10), vascular (n = 4), semantic (n = 2), and frontotemporal dementia (n = 1), progressive nonfluent aphasia (n = 1), corticobasal degeneration (n = 1), mixed dementia (Alzheimer's and cerebrovascular

disease; $n = 1$), and dementia with Lewy bodies ($n = 1$). The diagnosis was based on published criteria for these conditions (e.g., McKeith et al., 1996; McKhann et al., 1984; Neary et al., 1998).

All of the patients were bona fide—that is, there were no signs of negative response bias, suboptimal performance, or uncooperativeness in the clinical presentation and history. None was involved in litigation.

The second group consisted of 50 healthy, community-dwelling adults who were recruited as a control group for another study described in more detail by Wolf, Henry, Deike, Ebert, and Wallesch (2008). They received a sum of 20 euros (equivalent to about 30 US dollars) for their participation and signed an informed consent form. The controls were at least 50 years of age in order to be age matched to the mostly elderly population of neurological patients. The examination started with a clinical interview to exclude participants with neurological, psychiatric, or other diseases relevant for cognitive functioning. A total of 2 controls from originally 52 were excluded because they reported alcohol abuse. Mini-Mental State Examination (MMSE) scores (Folstein, Folstein, & McHugh, 1975) as well as Beck Depression Inventory (BDI) scores (Beck, Rush, Shaw, & Emery, 1979) were available for all controls. BDI scores varied from 0 to 15 with a mean score of 4.1 ($SD = 3.4$).

The major demographic characteristics of the two groups are summarized in Table 1. Post hoc comparisons showed that the patients were less educated, $t(113) = 4.48$, $p < .05$, and had lower MMSE scores, $t(51.9) = 5.01$, $p < .05$. Both groups were comparable in terms of age, $t(92.6) = 1.66$, *ns*, and gender, $\chi^2(1) = 3.05$, *ns*.

Procedure and instruments

An authorized German version of the NV-MSVT directions was used. Test presentation was done with the original computerized test version. The NV-MSVT consists of the following parts:

1. *List Presentation:* A total of 10 colored images are presented on the computer screen, each of them consisting of a pair of items that are closely associated in common human experience (e.g., scissors and paper). All 20 target items are to be named by the participants. The list is presented twice with a rate of one pair every four seconds.
2. *Immediate Recognition (IR) trial:* One part of the target item pairs is presented along with a foil not previously seen (e.g., scissors and elephant). Both items are to be named, and the target item that has been seen previously has to be chosen. This is done for all 20 target items, 1 at a time.
3. *Degraded Foil Sheet:* After the IR trial, a degraded foil sheet is presented to the participants. Degraded foils consist of 20 new items, which are degraded by a set of blank lines drawn through them. The participants do not have to name the items aloud, but have to carefully watch all the pictures for one minute. They are instructed to name them silently to themselves.

After a time lapse of nine minutes, the delayed recognition trial starts. Pairs of items are presented to the participants again. They are asked to choose the one picture of each pair that they have previously seen *on the computer*. This testing phase consists of the following three subtests, resulting in three separate parameters:

4. *Delayed Recognition (DR) trial:* The original target items from the list (e.g., scissors) are paired with foils from the degraded foil sheet (e.g., knife).
5. *Delayed Recognition–Variations (DRV):* The original target pairs (e.g., scissors and *yellow* paper) are presented together with almost identical, but slightly varied pictures (e.g., scissors and *white* paper). The original picture has to be identified.
6. *Delayed Recognition–Archetypes (DRA):* The foils from the IR trial (e.g., elephant) are presented with new images called *archetypes* by the author (e.g., black cat). The participants

TABLE 1
Demographic characteristics of the two groups

	Bona fide neurological patients (n = 65: 43 M/22 F)		*Healthy controls (n = 50: 25 M/25 F)*	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*
Age (years)	59.1 (16.1)	17–83	62.8 (7.1)	50–76
Education (years)	13.1 (3.4)	6.0–20.5	15.8 (3.0)	10–22.5
Mini-Mental State Examination[a]	25.7 (3.9)	15–30	28.7 (1.1)	26–30

Note. M = male. F = female.
[a]Mini-Mental scores were available for only 46 patients, but for all controls.

are required to select the foils from the IR trial. Naming the foils during the IR trial is essential to achieve this. According to preliminary studies done by Green (2008), even participants with intact memory are rarely able to recognize foils previously presented if they did not name them. The use of archetype pictures as new foils is intended to lower the difficulty of this subtest.

7. *Consistency (CNS):* Additionally, the level of consistency between DR and IR trials is calculated. Choices for a specific target item are being compared between IR and DR. To be consistent on one particular item, the examinee's response has to be right or wrong in both trials.
8. *Paired Associates (PA):* The participants are shown one part of the original target pairs (e.g., scissors); they are asked to recall what object went with it (e.g., paper).
9. *Free Recall (FR):* Finally, participants are required to freely recall all objects or animals that were contained in the original item list.

Standard administration procedures were followed except for two minor alterations (cf. Howe & Loring, 2009): Due to the severity of impairments in some patients and the low acquaintance of the mostly elderly participants with computer technology, the mouse was controlled by the examiner during the test administration. Thus, the examiner did not leave the room during the DR trial, contrary to the original test instructions. This alteration was mandated by the fact that patients suffering from dementia or other significant cognitive impairment generally experience serious problems using computer devices. Moreover, forgetting test instructions is a phenomenon frequently experienced with these patients. However, the NV-MSVT manual (Green, 2008) explicitly addresses this problem and mentions the possibility of adapting administration procedures for severely impaired patients. In order to ensure equal conditions for all participants, the modified testing procedure was used in all trials.

According to the classification rules recommended by the NV-MSVT, a case is classified as failing the test if: (A1) the mean of IR, DR, CNS, DRA, DRV, and PA is 90% or below or (A2) the mean of DR, CNS, DRA, and DRV is 88% or below. The profiles of individuals who score below one of these cutoffs are further analyzed in order to identify patients with authentic severe neurocognitive impairment (i.e., to reduce false-positive assignments). For being classified as suspect, the following criteria have to be met as well: (B1) PA must be fewer than 11 points lower than the average of DR, CNS, DRA, and DRV, and (B2) the mean of PA and FR (hard subtests) must be fewer than 20 points below the mean of IR, DR, and CNS (easy subtests), and (B3) the standard deviation of IR, DR, CNS, DRA, and DRV must be 12 or above. The classification rules are illustrated in Figure 1. For cases where one or more of the B criteria are not met although one of the A criteria would indicate poor effort, a dementia profile is assumed. This means that failure to meet any one of the three B criteria reduces prospects for suspect effort and makes genuine impairment more likely. These classification rules are described in detail in the test manual, but no information is available about how the rules and cutoff scores were established and optimized.

For the patient group, a number of neuropsychological test measures were available. For the present analyses, tests were chosen only if they had been given to at least 40 patients (i.e., more than 60% of the total patient group). In the following list, the number of available protocols is given in parentheses:

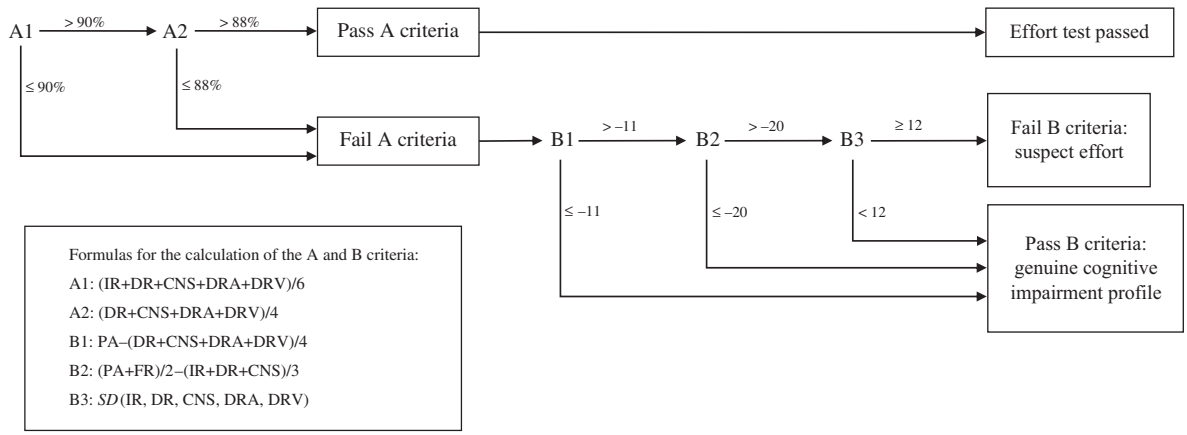

Figure 1. Flow chart depicting the two-step diagnostic procedure of the Nonverbal Medical Symptom Validity Test (NV-MSVT; A and B criteria).

1. The Mini-Mental State Examination (Folstein et al., 1975; $n = 46$).
2. A German-language short version of the Rey Auditory Verbal Learning Test (Rey, 1958; $n = 40$) for use in dementia assessment. A total of 10 words were tested over five trials: four consecutive learning trials and one free recall delayed by 15 minutes. Analysis was performed on the number of recalled items in the delayed trial and the average over the four consecutive learning trials.
3. A German-language 30-item adaptation of the Boston Naming Test (Merten, 2004; $n = 43$).
4. The subtests Digit Span Forward and Backward from the Wechsler Memory Scale–Revised (Wechsler, 1987; $n = 64$). In addition to the test scores, the Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994), which is a measure of low effort, was computed. A cutoff of 7/8 was used for classification.
5. The Corsi block tapping test. The score was calculated by counting the number of reliably tapped blocks (Milner, 1971; $n = 45$).
6. A figural fluency task developed by Regard (e.g., Regard, Strauss, & Knapp, 1982) and modified by Haid et al., 2002 ($n = 41$). The number of patterns correctly produced within three minutes was counted.
7. A test of motor reaction time to visual stimulation (alertness) from a comprehensive computerized test battery for attentional resources (TAP). Results were obtained by calculating the median of the reaction times (Zimmermann & Fimm, 2002; $n = 63$).
8. A test of selective attention (go/no-go) from the same test battery. Results were calculated using the median of the reaction times ($n = 62$).

Because of partly gross deviations from normal distributions and from homogeneity of variances, nonparametric statistics were exclusively employed for the group comparisons. In particular, H tests (Kruskal–Wallis) were used as an analysis of variance analogue for rank variables, and U tests (Mann–Whitney) were used for post hoc pairwise comparisons. For the latter tests, Bonferroni corrections for multiple comparisons were done.

RESULTS

The NV-MSVT results of the healthy controls and the patients are outlined in Table 2. The latter group was separated into patients with and without dementia. As can be seen from the table, the IR trial was the easiest subtest for all three groups. Both patients without dementia and healthy controls also scored high or very high on the subtests that followed, with the free recall task being the most difficult of the subtests. Table 2 also contains A1 scores (mean of all subtests except Free Recall) for all groups. For group comparison, separate Kruskal–Wallis tests (one-way analyses of variance

TABLE 2
Descriptive statistics of the Nonverbal Medical Symptom Validity Test

	Bona fide patients (n = 65)									*Kruskal–Wallis one-way ANOVA (H tests)*	*Pairwise differences (U tests)*[a]
	A: With dementia (n = 21)			*B: Without dementia (n = 44)*			*C: Healthy controls (n = 50)*				
	M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	χ^2	*Contrasts**
NV-MSVT IR	87.6	15.1	50–100	99.9	0.8	95–100	99.9	0.7	95–100	44.40*	A vs. B, C
NV-MSVT DR	85.7	14.6	55–100	95.5	6.7	70–100	96.9	5.2	75–100	11.54*	A vs. B, C
NV-MSVT CNS	80.5	18.4	35–100	95.3	6.9	70–100	96.8	5.2	75–100	16.32*	A vs. B, C
NV-MSVT DRV	79.5	17.5	50–100	93.7	8.7	70–100	96.2	5.3	80–100	20.03*	A vs. B, C
NV-MSVT DRA	85.2	12.7	65–100	93.0	7.6	65–100	96.2	6.4	65–100	17.88*	A, B vs. C
NV-MSVT PA	59.1	29.8	10–100	95.9	8.4	60–100	99.6	2.0	90–100	58.47*	A vs. B, C B vs. C
NV-MSVT FR	25.0	16.3	0–50	72.1	18.2	30–100	81.0	13.8	40–100	52.22*	A vs. B, C
Criterion A1	79.6	15.1	52–99	95.5	4.1	86–100	97.7	2.6	88–100	36.90*	A vs. B, C B vs. C

Note. NV-MSVT = Nonverbal Medical Symptom Validity Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; DRV = Delayed Recognition–Variations; DRA = Delayed Recognition–Archetypes; PA = Paired Associates; FR = Free Recall; A1 = mean score of IR, DR, CNS, DRV, DRA, and PA. ANOVA = analysis of variance.
[a]Corrected for ties.
*$p < .05$.

for nonparametric data) were performed. For all subtests, χ^2 values were significant, with Paired Associates and Free Recall showing the largest effects. Post hoc group comparisons revealed that patients with dementia scored significantly lower than healthy controls in all subtests. Delayed Recognition–Archetypes was the only subtest for which patients with dementia did not score significantly below patients without dementia. For one subtest, Paired Associations, all three groups scored significantly differently, with patients scoring lower than healthy controls.

The mean profiles of the three groups are illustrated in Figure 2. The groups appear to differ from each other in terms of the relative difficulty of some subtests. While Free Recall is the most difficult subtest for all groups, Paired Associates appears to be an easy task for healthy elderly people as well as for patients without dementia. In contrast, the mean Paired Associates score for patients with dementia amounts to less than 60%.

In the individual profile analyses, 20 patients (13 patients with dementia, 7 without dementia) and 1 control failed the test by failing the A criteria. When B criteria were added to the classification, none of the patients with dementia were finally classified as performing below their capacity. A total of 8 of these patients (36%) showed regular NV-MSVT results instead of a dementia profile. A total of 6 of the 44 patients without dementia (14%) had a NV-MSVT dementia profile. Also, the test scores of 1 of the healthy controls (2%) were classified as suggestive for a dementia profile. In fact, he showed no evidence of cognitive impairment but had an impulsive response style. Only 1 patient without dementia was classified as showing poor effort in the test. His clinical diagnosis was possible frontotemporal dementia (in an early stage, so that *DSM-IV* core criteria for dementia were not yet met). The profiles of these two cases are illustrated in Figure 3. A summary with the number of participants meeting the criteria can be found in Table 3.

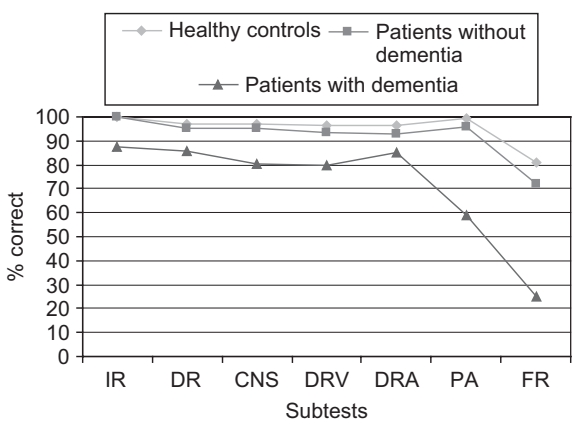

Figure 2. Test profiles of the three groups on the Nonverbal Medical Symptom Validity Test (mean scores). IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; DRV = Delayed Recognition–Variations; DRA = Delayed Recognition–Archetypes; PA = Paired Associates; FR = Free Recall.

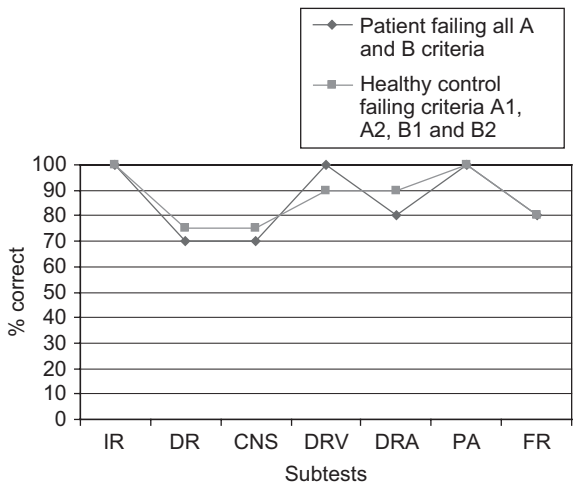

Figure 3. Two cases with atypical profiles. The patient was subsequently diagnosed as presenting suboptimal effort, the healthy participant as presenting a dementia profile. IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; DRV = Delayed Recognition–Variations; DRA = Delayed Recognition–Archetypes; PA = Paired Associates; FR = Free Recall.

TABLE 3
Number of cases meeting the different criteria for suspect effort

	Bona fide patients (n = 65)		
	With dementia (n = 21)	*Without dementia (n = 44)*	*Healthy controls (n = 50)*
Criterion A1	13	6	1
Criterion A2	10	5	1
Criterion A1 or A2	13	7	1
Criterion B1	1	6	1
Criterion B2	0	4	1
Criterion B3	4	1	0
Total classification			
Regular NV-MSVT results	8	37	49
Dementia profile	13	6	1
Suboptimal test motivation	0	1	0

Note. NV-MSVT = Nonverbal Medical Symptom Validity Test. While passing or failing A1 and A2 was based on the total sample, B1 to B3 were only checked if A1 or A2 was failed.

For the patient group, correlations of NV-MSVT variables with other neuropsychological performance tests were computed. The results (Table 4) show that NV-MSVT variables were associated with neuropsychological test scores in a differential way. The highest correlations amounted to .78 between Pair Associates and the Boston Naming Test and .75 between NV-MSVT Free Recall and Delayed Free Recall in a verbal learning task. While generally high correlations were found between NV-MSVT subtests and measures of verbal memory (with the Boston Naming Test representing a measure of semantic memory), low or zero correlations were found between NV-MSVT subtests and measures of attention (including Digit Span Forward and Backward). Also, the correlation between NV-MSVT scores and the Reliable Digit Span was nonsignificant for most subtests.

When using the Reliable Digit Span for detecting suboptimal effort, the original cutoff score would classify 12 out of 21 patients with dementia (57%) and 16 out of 43 patients without dementia (37%) as performing below their actual level of ability.

For the total group of patients, there was no strong link between general cognitive ability (as reflected by the MMSE scores) and passing or failing the A criteria of the NV-MSVT test. This is illustrated in Figure 4.

For the 50 healthy controls, no significant correlations were found between the Mini-Mental scores and NV-MSVT variables, with the exception of NV-MVST Free Recall ($r = .32$, $p < .05$). No significant correlations were found between Beck Depression Inventory scores and NV-MSVT variables.

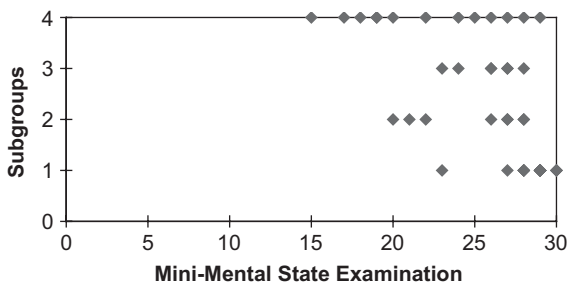

Figure 4. Association between Mini-Mental State scores and classification according to A criteria. Group 4: Patients with dementia failing A criteria (Nonverbal Medical Symptom Validity Test; NV-MSVT dementia profile); Group 3: Patients without dementia failing A criteria (NV-MSVT dementia profile); Group 2: Patients with dementia passing NV-MSVT A criteria; Group 1: Patients without dementia passing NV-MSVT A criteria.

DISCUSSION

This study is one of the first to present empirical data for a new symptom validity test, the Nonverbal Medical Symptom Validity Test (NV-MSVT). The study primarily addressed the question of

TABLE 4
Correlations between the Nonverbal Medical Symptom Validity Test variables and other neuropsychological tests, for the patient group

	Nonverbal Medical Symptom Validity Test						
	IR	*DR*	*CNS*	*DRV*	*DRA*	*PA*	*FR*
Age (years)	−.30*	−.43*	−.43*	−.37*	−.40*	−.47*	−.54*
Education (years)	.07	.24	.19	.09	.17	.16	.08
Mini-Mental State	.58*	.65*	.61*	.47*	.51*	.60*	.63*
AVLT Short Version, Verbal Learning	.55*	.54*	.53*	.35*	.37*	.66*	.65*
AVLT Short Version, Delayed Recall	.60*	.42*	.53*	.33*	.45*	.65*	.75*
Boston Naming Test	.59*	.59*	.64*	.51*	.55*	.78*	.66*
Figural Fluency	.34*	.32*	.35*	.42*	.39*	.51*	.62*
WMS–R Digit Span Forward	.15	.28*	.22	.12	.16	.28*	.34*
WMS–R Digit Span Backward	.09	.34*	.22	.14	.13	.17	.30*
Corsi Block Tapping	.47*	.39*	.45*	.14	.36*	.37*	.38*
Motor Reaction Time (Alertness)	.13	.06	.12	−.05	.03	−.11	−.23
Go/No-go, Reaction Time	−.03	−.10	−.07	−.20	.00	−.27*	−.23
Go/No-go, Number of Errors	−.10	−.26*	−.19	−.11	−.30*	−.22	−.32*
Reliable Digit Span	.13	.31*	.26	.05	.14	.21	.28*

Note. Total $n = 65$. IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; DRA = Delayed Recognition–Archetypes; DRV = Delayed Recognition–Variations; PA = Paired Associates; FR = Free Recall; AVLT = Auditive Verbal Learning Test; WMS–R = Wechsler Memory Scale–Revised.
*$p < .05$.

specificity. The specificity of an SVT is of critical importance, especially with regard to those patients suffering from significant cognitive impairment (e.g., dementia). Although it is easy for many groups of neurological patients to meet the task demands of most SVTs, the same is not true for patients suffering from dementia. Thus, in a study by Merten et al. (2007), most of the patients with early Alzheimer's disease failed to reach scores above empirical cutoffs in SVTs such as the Word Memory Test and the Amsterdam Short-Term Memory Test, but the same was also true for Reliable Digit Span results. It would, of course, be a major diagnostic error to infer suboptimal effort (or, even worse, malingering) based on such results. Scores beneath empirically established cutoffs in SVTs usually raise concerns about the validity of neuropsychological test results; however, the examiner is often left with the question of whether true cognitive impairment of some severity could have accounted for this failure. Thus, it is of major importance to scrutinize the classification accuracy of every SVT in patients with moderate and severe cognitive impairment.

The NV-MSVT uses a two-step diagnostic procedure to discriminate authentic from nonauthentic test profiles: In a first step test scores are checked against empirical cutoffs (A criteria); then, if a person scores below that cutoff, it has to be examined in a second step whether the test profile reflects genuine cognitive impairment or rather resembles the profile of individuals with negative response bias (B criteria). In this study, 20 of 65 neurological patients failed the A criteria. Taking all three B criteria into account, a single patient remained whose scoring profile had to be classified as nonauthentic. Thus, a false-positive rate of 1.5% was obtained for the patient group. This rate, however, can be assumed only if the patient in question was really performing at the level of his abilities. Also, one member of the control group met the A criteria for low effort. While the criteria B1 and B2 would have indicated a nonauthentic profile in this person, B3 did not. Thus, in the final classification, his scores would not be classified as a low-effort profile but as a dementia profile. Hence, his classification was wrong. While the approach of the hierarchical NV-MSVT analysis is based upon the principle that dementia needs to be ruled out before poor effort is concluded, dementia may, of course, not be assumed if a person does not present clinical features compatible with that diagnosis even if NV-MSVT results would suggest a "dementia profile." Thus, in a recent pilot study done by Brockhaus, Kok, and Witte (2008), 11 geriatric health workers tried to fake dementia in a simulation design. In fact, the Word Memory Test profile analysis correctly classified 6 of them as presenting suboptimal effort, but 5 obtained a WMT dementia profile.

It is still not clear from the interpretation guidelines found in test manuals how such cases are to be dealt with and to which degree clinical judgment and collateral information, such as information about clinical presentation and everyday functioning, are to be taken into account in such cases. Clear decision rules would be desirable.

Curiously enough, there were parallels between the member of the control group failing the A criteria and the one patient who failed all A and B criteria and who had the diagnosis of possible frontotemporal dementia: Both demonstrated an impulsive style of selecting test responses in the recognition trials. As a consequence, they obtained comparatively low scores in relatively easy subtests (DR, CNS, DRA), but they did very well on harder subtests (PA and FR), so their failure has rather to be attributed to an impulsive response style than to negative response bias (cf. Figure 3). Though there has been some informal discussion for many years among experts about whether impulsive behavior may lead to SVT failure, empirical evidence is still lacking to substantiate this hypothesis. In this context, a further examination of symptom validity indicators in patients with frontotemporal dementia may be useful.

It is not only dementia that may account for the significant difficulties that authentic patients may have with symptom validity tests. Patients with aphasia, spatial neglect, amnestic syndromes, confusional states, or frontal lobe syndromes may experience similar difficulties. Thus, the denomination *dementia profile* may perhaps be changed into *profile of genuine cognitive impairment*. While there is casual evidence that some of the patients with the conditions exemplified above have no difficulties with particular tests, this, of course, does not mean that patients with moderate to severe neurocognitive dysfunctions other than dementia generally have no difficulties. Like any other test, NV-MSVT test performance requires basic cognitive functioning. It is correlated with measures of episodic and semantic memory, figural fluency, and spatial span as can be shown by the correlational patterns obtained for the patient sample (Table 4). Similar analyses have previously been performed by Merten et al. (2007) for different symptom validity measures.

The low or even nonsignificant correlations obtained between NV-MSVT parameters and Reliable Digit Span scores replicate the results of a number of previous studies that found that pairs of symptom validity measures often intercorrelated modestly, and some correlations were not even

statistically significant (e.g., Inman & Berry, 2002; Merten et al., 2007; Nelson et al., 2003). The consequences of such findings for the validity of SVTs are not yet clear.

The widely missing significant correlations between NV-MSVT scores and MMSE or BDI scores in the group of healthy elderly people have to be viewed with some circumspection. There was very little variation in Mini-Mental scores in the elderly group, and though there was some report of depressive symptoms, the presence of clinically relevant depression was among the exclusion criteria for the control group. Thus, the nonsignificant correlations between BDI scores and NV-MSVT subtests cannot be generalized to populations with clinically relevant depression.

Results of controls and neurological patients without dementia demonstrate that the NV-MSVT is a fairly easy test: With the exception of Free Recall, the two groups performed at ceiling level in all subtests. In contrast, patients with dementia obtained significantly lower scores in the easy subtests (IR, DR, DRA, and DRV), with this difference increasing in the harder tests (PA and FR). A total of 13 of 20 patients meeting the A criteria for low effort also met DSM-IV criteria for dementia. The remaining 7 of those 20 patients also suffered from degenerative diseases mostly associated with dementia (e.g., Huntington's disease, progressive supranuclear palsy, or normal pressure hydrocephalus) but did not fulfill general dementia criteria at the time of testing.

These results support the notion that using empirically established cutoff scores in isolation may lead to an unacceptably high rate of false positives in patients with severe cognitive impairment (Merten et al., 2007; Teichner & Wagner, 2004). The use of a profile analysis in SVTs might be an option to circumvent this major problem. There is some evidence that patients suffering from dementia show a different profile in SVTs (like WMT or NV-MSVT) than simulators or patients with other diagnoses (Gill, Green, Flaro, & Pucci, 2007; Howe, Anderson, Kaufman, Sachs, & Loring, 2007; Richman et al., 2006). While there is some evidence from the WMT and MSVT that scores vary depending on state of progression (beginning vs. advanced dementia), there has been no study so far addressing the presence of distinguishable profiles in different types of dementia. The sample used in the present study is too small to analyze different subgroups of dementing conditions.

In conclusion, the results of the present study support the data published by Green (2008): Due to its two-step diagnostic procedure, the NV-MSVT has a high specificity for suboptimal effort in neurological patients and healthy controls. In our sample, a single patient was misdiagnosed resulting in a specificity of 98.5% for neurological patients and 100% for controls. However, the study does not allow conclusions about the sensitivity of the NV-MSVT. Thus, it is still to be demonstrated by independent research that the NV-MSVT is an instrument appropriate for the use in the clinical and forensic context.

Original manuscript received 12 December 2008
Revised manuscript accepted 31 January 2009
First published online day month year

REFERENCES

American Psychiatric Association. (1995). *Diagnostic and statistical manual of mental disorders: DSM-IV:* International version with ICD-10 codes. Washington, DC: Author.

Batt, K., Shores, A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society*, *14*, 1074–1080.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.

Boone, K. B. (Ed.). (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford.

Brockhaus, R., Kok, P., & Witte, N. (2008, October). *Der Word Memory Test identifizierte Experten, die eine Demenz simulierten* [The Word Memory Test identified experts simulating dementia]. Poster presented at the 23rd Annual Convention of the German Neuropsychological Society (GNP), Tuebingen, Germany.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, *20*, 419–426.

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 Psychiatry Research*, *12*, 189–198.

Gill, D., Green, P., Flaro, L., & Pucci, T. (2007). The role of effort testing in independent medical examinations. *The Medico-Legal Journal*, *75*, 64–72.

Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research*, *78*, 199–208.

Graue, L. O., Berry, D. T. R., Clark, J. A., Sollmann, M. J., Cardi, M., Hopkins, J., et al. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings. *The Clinical Neuropsychologist*, *21*, 929–942.

Green, P. (2003). *Green's Word Memory Test. User's manual*. Edmonton, Canada: Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows. User's manual*. Edmonton, Canada: Green's Publishing.

Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test (NV-MSVT) for Microsoft Windows. User's Manual 1.0*. Edmonton, Canada: Green's Publishing.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218–224.

Haid, T., Martl, C., Schubert, F., Wenzl, M., Kofler, M., & Saltuari, L. (2002). Der "HAMASCH 5 Punkt Test"—erste Normierungsergebnisse [The HAMASCH Five-Point Test: First results of a normative study] [Abstract]. *Zeitschrift für Neuropsychologie*, *13*, 233.

Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. M., & Brennan, A. (2005). WAIS Digit Span-based indicators of malingered neurocognitive dysfunction: Classification accuracy in traumatic brain injury. *Assessment*, *12*, 429–444.

Howe, L. L. S., Anderson, A. M., Kaufman, D. A. S., Sachs, B. C., & Loring, D. W. (2007). Characterization of the Medical Symptom Validity Test in evaluation of clinically referred memory disorders clinic patients. *Archives of Clinical Neuropsychology*, *22*, 753–761.

Howe, L. L. S., & Loring, D. W. (2009). Classification accuracy and predictive ability of the Medical Symptom Validity Test's dementia profile and general memory impairment profile. *Archives of Clinical Neuropsychology*, *23*, 329–342.

Inman, T. H., & Berry, D. T. R. (2002). Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology*, *17*, 1–23.

Larrabee, G. J. (Ed.). (2007). *Assessment of malingered neuropsychological deficits*. Oxford, UK: Oxford University Press.

McKeith, I. G., Galasko, D., Kosaka, K., Perry, E. K., Dickson, D. W., Hansen, L. A., et al. (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): Report of the Consortium on DLB International Workshop. *Neurology*, *47*, 1113–1124.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Merten, T. (2004). Entwicklung einer deutschen Kurzform des Boston Naming Test [Development of a German short version of the Boston Naming Test]. *Neurologie & Rehabilitation*, *10*, 281–287.

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*, 308–318.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*, 272–277.

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., et al. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*, 1546–1554.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist*, *17*, 263–272.

Regard, M., Strauss, E., & Knapp, P. (1982). Children's production of verbal and nonverbal fluency tasks. *Perception and Motor Skills*, *55*, 839–844.

Rey, A. (1958). *L'examen clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de Paris.

Richman, J., Green, P., Gervais, R., Flaro, L., Merten, T., Brockhaus, R., et al. (2006). Objective tests of symptom exaggeration in independent medical examinations. *Journal of Occupational & Environmental Medicine*, *48*, 303–311.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, *19*, 455–464.

Turner, T. H., Horner, M. D., Edmiston, K., & Bachman, D. L. (2008). Insufficient effort or severe impairment? Utility of the Test of Memory Malingering (TOMM) in a memory disorder clinic [Abstract]. *Journal of the International Neuropsychological Society*, *14(*Suppl. 1), 64.

Wechsler, D. (1987). *Wechsler Memory Scale–Revised. Manual*. San Antonio, TX: Psychological Corporation.

Weinborn, M., Woods, S. P., & Fox, A. (2008, July). *A validation of the Verbal and Nonverbal Medical Symptom Validity Tests using a simulation design*. Poster presented at the meeting of the International Neuropsychological Society, Buenos Aires, Argentina.

Wolf, S. A., Henry, M., Deike, R., Ebert, A. D., & Wallesch, C. W. (2008). Verdachtsdiagnose Alzheimer-Demenz. Bei welchen Patienten erfolgt eine neuropsychologische Abklärung? [Suspected Alzheimer's disease. Selection of outpatients for neuropsychological assessment]. *Nervenarzt*, *79*, 444–453.

Zimmermann, P., & Fimm, B. (2002). *Testbatterie zur Aufmerksamkeitsprüfung (TAP) Version 1.7* [Test for Attentional Performance (TAP) Version 1.7]. Herzogenrath, Germany: Psytest.