Archives of Clinical Neuropsychology
20 (2005) 719–726

Archives of CLINICAL NEUROPSYCHOLOGY

Analog validation of German-language symptom validity tests and the influence of coaching ☆

Thomas Merten [a,*], Paul Green [b], Matthias Henry [c], Nina Blaskewitz [c], Robbi Brockhaus [d]

[a] *Vivantes Netzwerk für Gesundheit, Klinikum im Friedrichshain, Klinik für Neurologie, Landsberger Allee 49, D-10249 Berlin, Germany*
[b] *Neurobehavioral Associates, Edmonton, Canada*
[c] *Humboldt University, Berlin, Germany*
[d] *Alexianer Krankenhaus, Krefeld, Germany*

Accepted 9 April 2005

Abstract

Although symptom validity testing is an integral part of the repertory of neuropsychologists in a number of countries, this is not yet true for Germany. The German adaptations of two effort tests, the Medical Symptom Validity Test (MSVT) by Green and the Amsterdam Short-Term Memory Test (ASTM) by Schmand et al., were investigated with a German-language sample. An analog study was performed with 18 healthy experimental malingerers and 18 controls with a mean age of 25.4 years. The scenario contained detailed information about mild post-traumatic cognitive impairment, as well as an explicit warning against symptom exaggeration. In addition to MSVT and ASTM, the Trail Making Test (TMT), the Complex Figure Test (CFT), and Digit Span were performed. Half of the sample were also given Rey's 15-Item-Test (FIT). Both groups were significantly different in all effort and performance measures, with the exception of the ratio TMT-B:TMT-A. With MSVT and ASTM, correct classification of group membership was between 97 and 100%. For the ratio TMT-B:TMT-A, there was a considerable overlap in the test scores for the two groups and the sensitivity of the FIT was too low. Although the ASTM and the MSVT were identified by a number

☆ The results of this study were presented as a poster at the 19th Annual Meeting of the German Neuropsychological Society (GNP) in Munich, September 2004. A more detailed report on the study as well as a translation of the scenario and the questionnaires used can be obtained from the authors.

* Corresponding author. Tel.: +49 3042211663.
E-mail address: thomas.merten@vivantes.de (T. Merten).

0887-6177/$ – see front matter
doi:10.1016/j.acn.2005.04.004

of subjects as possible effort measures, both tests obtained very good results within this analog design.

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

1. Introduction

Psychologists have always made comments in reports about the effort that patients apply to testing, acknowledging that it is a significant variable, but objective testing of effort is a recent development. It is now standard practice in the USA and Canada for neuropsychologists to use effort testing routinely, especially if financial compensation is an issue but also more generally because poor effort will usually go undetected if it is not measured. In contrast to the remarkable interest that malingering research has received in North America, effort testing has had little impact so far on European countries. A major exception is the Amsterdam Short-Term Memory Test (Schagen, Schmand, de Sterke, & Lindeboom, 1997; Schmand et al., 1998), which has been adapted into English and German from the original Dutch. Also, the oral and computerized versions of Green's Word Memory Test (WMT; Green, 2003) are available in German, Spanish, Dutch, French, English, Danish, and Turkish.

In German-speaking countries, until now there have been a very small number of relevant publications on the subject of suboptimal performance. Also, symptom validity tests (SVTs) have not yet been employed by clinicians on a widespread basis. There have been two German-language studies (Brockhaus & Merten, 2004) using the WMT, one with patients with mental retardation, and the other with experimental malingerers and normal controls. The first study showed that the WMT was able to identify high performance motivation in almost all the patients with mental retardation. Only one participant was incorrectly classified. The second study compared performance of a healthy group of experimental malingerers ($n = 100$) with a healthy control group after standard test instructions ($n = 27$). The WMT was able to classify cases of good effort versus cognitive exaggeration with 100% accuracy in the study, replicating the finding of 100% correct classification by the WMT in an independent simulator study in English (Tan, Slick, Strauss, & Hultsch, 2002). In a more recent study, which is one of the few empirical studies on symptom validity testing ever published in the German language, Merten, Henry, and Hilsabeck (2004) used an analog design with a small sample size to investigate the applicability of German adaptations of foreign-language tests. In that study, three SVTs were used: the Word Memory Test, the Amsterdam Short-Term Memory Test, and the Word Completion Memory Test. For the SVTs, considerable effect sizes and a correct group classification of 95–100% was obtained. Although ratings of test difficulty given by all participants revealed that the SVT tasks were judged to be easier than the performance tests, analogue malingerers were unable to simulate the response patterns of truly brain-injured patients on the SVT measures.

The present study was conceived to follow that line of research with a different set of tests and a different scenario, which included an elaborate warning that effort measurement was to be part of the assessment procedure and that overplaying the role would result in a loss of credibility in the overall pattern of performance. Moreover, a new test, the MSVT (Medical Symptom Validity Test; Green, 2004) was to be tested in the context of an analog design. The main question of the study was how well German test adaptations would work.

One additional question to be addressed by the study was how well the participants would be able to identify the real symptom validity tests among a set of instruments, after receiving explicit warnings about effort assessment. To this effect, another test, Rey's 15-Item-Test (FIT; Rey, 1958), was given to half of the experimental group and to half of the controls to see whether this test would be recognized as an effort measure very easily and, thereby, deflect attention from the other SVTs. After the experiment, the subjects were asked to identify possible effort measures. The hypothesis was that those participants who were given the FIT would easily recognize it as an SVT and, as a consequence, identify the other effort tests with less frequency.

2. Method

2.1. Research participants

Thirty-six healthy, well-educated fluent speakers of German participated in the experiment. The group consisted of 12 males and 24 females with a mean age of 25.4 years (S.D. = 4.0; range: 20–35 years) and a mean of 17.4 years of education (S.D. = 2.5). Age and education were not significantly different for the experimental and the control group (age: 25.3 vs. 25.4 years; education: 17.5. vs. 17.4 years). Twenty-six participants (72%) were university undergraduates while five held a university degree (14%). Four non-native speakers with a fluent command of German, who were all university undergraduates, were included.

2.2. Procedure

Two experimental conditions were investigated on the basis of a pre-established randomized plan: group membership (experimental malingerers vs. full-effort participants) and inclusion of Rey's FIT (yes vs. no). The scenario for the experimental malingerers was designed to acquaint the participants with the role of a person who had suffered a car accident 2 years before and was preparing for a neuropsychological assessment. After the accident, the person had suffered from mild cognitive impairment, which had gradually improved with time. The person was asked to imagine that, as the legal procedure to gain compensation had been delayed for a long time, the symptoms had improved considerably. However, the litigant would feel entitled to receive compensation for the handicaps and the social impediments suffered in the months following the accident. Therefore, he or she would decide to pretend that the symptoms had persisted and that they were still significant. The scenario also contained a strong element of coaching. It was to be assumed that the litigant's attorney had recommended not overdoing symptom presentation or else the litigant might fail testing done by an experienced neuropsychologist using effort tests.

To ensure role understanding, a 10-item multiple-choice questionnaire was developed. Apart from ensuring role commitment, the questionnaire was considered to be part of the coaching procedure. An additional element of warning was included in the instructions given to the experimental malingers at the beginning of the test session proper. They were told that the evaluation would include measures of their test motivation. In contrast, the controls did not receive such detailed coaching and they were just given brief instructions to put forth a full effort in the testing session.

To check role commitment after the test session, a separate set of questions was answered. Finally, both the members of the experimental group and the controls were asked whether, in their opinion, there had been tests included which might have exclusively served to detect symptom exaggeration or malingering.

2.3. Instruments

The following instruments were given individually in the order indicated here:

(1) The Medical Symptom Validity Test (MSVT; Green, 2004) Learning Trials and Immediate Recognition (IR). In the oral form of this test, used in the current study, the examiner reads aloud ten word pairs, such as “soccer ball” and “jet plane”. Immediate Recognition consists of choosing one of the words in the list (e.g., “soccer”) when given a pair of words containing one list word and one non-list, foil word (e.g., “soccer” and “basketball”).
(2) The Trail Making Test (TMT; Reitan, 1992). As a malingering measure, the ratio TMT-B:TMT-A was computed using a cutoff, 1.49/1.50 which corresponds to the 5th percentile of the total clinical sample studied by Iverson, Lange, Green, and Franzen (2002).
(3) The 15-Item Test (Rey, 1958) using a cutoff of 7. As noted above, this test was given to only 50% of the participants.
(4) The copy trial of Rey’s Complex Figure Test (CFT; Osterrieth, 1945).
(5) The MSVT, Delayed Recognition (DR), Paired Associates, and Free Recall trials. The same type of recognition task as described at (1), but using different foil words, is performed after a 10-min delay (DR). Immediately after that, in the Paired Associates subtest, the person is told the first word and is asked to say the second word (e.g., “soccer”, to which the correct response would be “ball”). Finally, in the Free Recall subtest, the person is asked to say as many words as possible from the list, in pairs or singly. A consistency score from the IR and DR trials is computed.
(6) The WAIS-R subtests, Digit Span forward and backward (Wechsler, 1981). In addition to the test scores, the Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994) was used as a further measure of low effort.
(7) The Amsterdam Short-Term Memory Test (Schmand et al., 1998).
(8) The 30 min recall trial of the CFT.

3. Results

The results obtained by the two groups are presented in Tables 1 and 2. All cutoffs employed in this study are listed in Table 2. Effect sizes for the symptom validity tests, the Reliable

Table 1
Raw scores obtained by two groups for all measures

Instrument	Controls ($n = 18$)[a]		Experimental malingerers ($n = 18$)[a]		U test (Z)
	m	S.D.	m	S.D.	
Symptom validity tests					
ASTM (raw scores)	89.4	0.8	64.2	8.9	−5.20*
MSVT—IR (%)	99.0	3.5	65.0	12.0	−5.42*
MSVT—DR (%)	99.5	1.5	61.0	16.0	−5.38*
MSVT—C (%)	99.0	2.5	62.0	14.5	−5.38*
FIT (raw scores)	15.0	0	11.2	2.7	−3.49*
Symptom validity parameters derived from other tests					
RDS (raw scores)	10.4	1.9	5.4	2.1	−4.72*
TMT-B:TMT-A (ratio)	2.4	0.9	2.0	0.8	−1.64
Neuropsychological tests					
MSVT P. Associates (%)	96.0	6.0	45.0	13.0	−5.25*
MSVT Free Recall (%)	91.0	3.0	38.0	14.5	−5.15*
CFT—IR (raw scores)	27.8	2.5	21.8	3.9	−4.20*
CFT—DR (raw scores)	19.7	4.7	11.0	4.8	−4.21*
TMT-A (s)	23.7	10.8	57.7	29.6	−4.56*
TMT-B (s)	51.4	13.3	129.1	121.8	−3.55*
Digit Span forward	8.7	2.2	3.4	1.8	−4.77*
Digit Span backward	7.2	1.9	3.6	1.5	−4.33*

Notes. ASTM, Amsterdam Short-Term Memory Test score out of 90; MSVT, Medical Symptom Validity Test, in percent correct; FIT, 15-Item Test score out of 15; CFT, Complex Figure Test; TMT, Trail Making Test; RDS, Reliable Digit Span as sum of reliable span forward and backward; MSVT—IR, Immediate Recognition; MSVT—DR, Delayed Recognition; MSVT—C, Consistency; CFT—IR, Complex Figure Test Immediate Recall; CFT—DR, Delayed Recall.

[a] Except for FIT ($n = 9$ for each group).

* $p < .05$.

Digit Span, and the performance measures were considerable. The exception was the ratio TMT-B:TMT-A, which did not significantly differentiate between analog malingerers and controls.

For both the MSVT and the ASTM, group assignment was perfect or close to perfect. Only one person in the control group scored at the proposed cutoff of 17/20 or 85% correct on MSVT—Immediate Recognition. In contrast to these two tests, results for the RDS were less favorable. The total accuracy of classification for both FIT and the ratio TMT-B:TMT-A was in the chance range (56 and 50%, respectively).

The participants' answers to the question about which instruments they thought might be measuring effort were analyzed. The ASTM, TMT-A, and MSVT were the tests most often suspected of being effort tests (experimental group: 4, 5, and 5, respectively; controls: 4, 2, and 2). Digits Forward was named three times, TMT-B twice, and Digits Backward was named once by experimental malingerers. One control named the copy trial of the CFT as a suspected effort test. The FIT and CFT, delayed recall, were not so named by any of the participants.

Table 2
Observed variation and quality of classification for symptom validity tests and indicators

		Controls				Experimental malingerers				
Instrument	Cutoff	Range	Hits	Misses	Specificity (%)	Range	Hits	Misses	Sensitivity (%)	Correct classification (total) (%)
ASTM	84/85	88–90	18	0	100	50–81	18	0	100	100
MSVT—IR	17/18	17–20	17	1	94	8–16	18	0	100	97
MSVT—DR	17/18	19–20	18	0	100	7–17	18	0	100	100
MSVT—C	17/18	18–20	18	0	100	8–17	18	0	100	100
FIT	7/8	15	9	0	100	6–15	1	8	11	56
RDS	7/8	6–13	16	2	89	2–8	15	3	83	86
TMT-A:TMT-B	1.49/1.50	0.9–4.4	14	4	78	1.0–4.9	4	14	22	50

Note. ASTM, Amsterdam Short-Term Memory Test; MSVT, Medical Symptom Validity Test; IR, Immediate Recognition; DR, Delayed Recognition; C, Consistency, FIT, 15-Item Test; RDS, Reliable Digit Span, TMT, Trail Making Test.

4. Discussion

Two new German-language adaptations of symptom validity tests were investigated. As very few SVTs are available to neuropsychologists in Germany, the present results may be viewed as being of considerable practical importance for this country. The results suggest that the German adaptations work as predicted in an analog design. Very powerful group differences were observed on several effort tests (ASTM, MSVT subtests and the RDS) between those who were asked to try their best and those who were asked to fake cognitive impairment. On the ASTM, there was no overlap in scores between the good effort group (mean score: 89.4; S.D. = 0.8) versus the cognitive exaggeration group (mean score: 64.2; S.D. = 8.9). Similarly, there was no overlap between the two groups in their scores on MSVT consistency or on the MSVT Delayed Recognition, Paired Associates and Free Recall scores. Using the German ASTM scores or the German MSVT subtest scores, we could very accurately classify people as belonging to the good effort group or the cognitive exaggeration group. On the other hand, neither the original Rey's FIT nor the TMT ratio performed any better than chance.

Contrary to expectations, the FIT was not perceived to be an effort test. None of the subjects named this test as a potential candidate for malingering detection, whereas a number of them thought the TMT-A, ASTM, or MSVT might be effort measures. There was a clear tendency among coached participants to be more aware of potential effort measures. However, it is interesting that the correct identification of the ASTM and the MSVT as effort tests did not prevent participants from scoring below the cutoffs on these tests.

The coaching procedure employed in this study can best be described as "detailed information plus warning". From coaching studies currently available it has become clear that the exact nature and the amount of coaching vary considerably. This may be important for a comparative evaluation of the results (e.g., Dunn, Shear, Howe, & Ris, 2003). Differential effects of various coaching strategies are currently being investigated in an ongoing study by the first author and his group.

A weakness of any simulator study is that people who are asked to exaggerate cognitive deficits in an experiment, even when playing a role as part of a well-described scenario involving a financial claim, do not actually have such a claim and, therefore, might not behave exactly as do actual patients. It has also become clear that simulator studies in which true patients assume the role of experimental malingerers can be considered superior (Ju & Varney, 2000). A more important limitation, however, is that in the context of this study simulated poor performance could not be compared to the scores from people with true clinical impairment. A study like this can only be considered as a single step in the complex process of test development.

Another potentially limiting factor in the current study is that, although they scored almost 100% correct on the ASTM and MSVT effort subtests, showing that the tests are extremely easy, the volunteers were highly educated. Some patients might not be able to do that well and so clinical studies are needed.

Based on the current study, the MSVT, the ASTM, and Reliable Digit Span all show potential as good effort tests, which could be used with German patients. However, much more research on effort testing is needed in Europe to come close to the neuropsychological testing standards, which have become customary in Canada and the United States.

References

Brockhaus, R., & Merten, T. (2004). Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test [Neuropsychological assessment of suboptimal performance: The Word Memory Test]. *Nervenarzt*, *75*, 882–887.

Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology*, *18*, 121–134.

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.

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

Iverson, G. L., Lange, R. T., Green, P., & Franzen, M. D. (2002). Detecting exaggeration and malingering with the Trail Making Test. *The Clinical Neuropsychologist*, *16*, 398–406.

Ju, D., & Varney, N. R. (2000). Can head injury patients simulate malingering? *Applied Neuropsychology*, *7*, 201–207.

Merten, T., Henry, M., & Hilsabeck, R. (2004). Symptomvalidierungstests in der neuropsychologischen Diagnostik: Eine Analogstudie [The use of symptom validity tests in neuropsychological assessment: An analog study]. *Zeitschrift für Neuropsychologie*, *15*, 81–90.

Osterrieth, P.-A. (1945). Le test de copie d'une figure complexe. Contribution à l'étude de la perception et de la mémoire [The Complex Figure Copying Test. Contributions to the study of perception and memory]. *Archives de Psychologie*, *30*, 205–353.

Reitan, R. M. (1992). *Trail Making Test. Manual for administration and scoring*. South Tucson, AZ: Reitan Neuropsychology Laboratory.

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

Schagen, S., Schmand, B., de Sterke, S., & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology*, *19*, 43–51.

Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H. L. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery, and Psychiatry*, *64*, 339–343.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, *16*, 495–505.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. San Antonio, TX: Psychological Corp.