



## Swiss population-based reference data for six symptom validity tests<sup>1</sup>

Peter Giger<sup>a\*</sup> and Thomas Merten<sup>b</sup>

<sup>a</sup>University of Berne, Switzerland

<sup>b</sup>Vivantes Klinikum im Friedrichshain, Berlin, Germany

### ARTICLE INFORMATION

Manuscript received: 24/08/2013

Revision received: 01/09/2013

Accepted: 15/09/2013

#### Keywords:

Symptom validity testing

Reference data

Neuropsychological assessment

Negative response bias

### ABSTRACT

Symptom validity test (SVT) results should be resistant against sociodemographic variables. Healthy, cooperative respondents should be able to pass these tests. The purpose of the study was to collect reference data for a selection of SVTs (Medical Symptom Validity Test, Structured Inventory of Malingered Symptomatology, Amsterdam Short-Term Memory Test, Emotional Numbing Test, Reliable Digit Span, Maximum Span Forward). A representative population-based sample of 100 German speaking Swiss citizens from 18 to 60 years of age was investigated. Multiple regression analyses revealed that age and verbal intelligence had an effect on various SVTs, whereas sex and education did not. The rate of positive test scores ranged from 1% (Emotional Numbing Test, Structured Inventory of Malingered Symptomatology) to 4% (Maximum Span Forward). One of the pertinent questions is if such positive results in reference or normative samples represent false positives or true positives and how to deal with the problem.

© 2013 Colegio Oficial de Psicólogos de Madrid. All rights reserved.

### Datos normativos para la población suiza de seis pruebas para evaluar la validez de los síntomas

#### RESUMEN

Los resultados encontrados en la evaluación de la validez de los síntomas (SVT) deberían ser inmunes a las variables sociodemográficas. Así, los sujetos sanos y cooperadores deberían superar los tests de SVT. El objetivo de este estudio es obtener datos de referencia de una serie de pruebas utilizadas en la evaluación de la validez de los síntomas, en una muestra representativa de 100 ciudadanos suizos, germano-hablantes, de entre 18 y 60 años: Medical Symptom Validity Test, Structured Inventory of Malingered Symptomatology, Amsterdam Short-Term Memory Test, Emotional Numbing Test, Reliable Digit Span y Maximum Span Forward. Los análisis de regresión múltiple reflejaron que las variables edad e inteligencia verbal afectaron a los resultados de varias de las pruebas, mientras que no fue así para las variables sexo y nivel educativo. La tasa de resultados positivos osciló entre el 1% (Emotional Numbing Test, Structured Inventory of Malingered Symptomatology) y el 4% (Maximum Span Forward). Una cuestión relevante que se desprende de este estudio es si dichos resultados positivos en las muestras de referencia o normativas representan sujetos falsos positivos o verdaderos positivos y cómo abordar este problema.

© 2013 Colegio Oficial de Psicólogos de Madrid. Todos los derechos reservados.

#### Palabras clave:

Validez de los síntomas

Datos normativos

Evaluación neuropsicológica

Sesgo de respuesta negativo

In forensic neuropsychological practice, symptom validity assessment has become common practice (Heilbronner et al., 2009; Sweet & Guidotti-Bretting, 2013) so much that test profiles without appropriate validity check are considered to be incomplete and potentially uninterpretable. In recent years, symptom validity assessment has attracted increasing interest in clinical and rehabilitative contexts too (e.g., Carone, Bush, & Iverson, 2013; Göbber, Petermann, Piegza, & Kobelt, 2012). Negative response bias

can manifest itself in two different ways (e.g., Dandachi-Fitzgerald & Merckelbach, 2013): 1) as false, distorted symptom report (usually in the form of symptom over-reporting); or 2) as distorted behavioral presentation of symptoms or as impairment (such as underperformance in cognitive assessment).

Both manifestations can occur in combination or by themselves. Accordingly, detection strategies focus on either underperformance or over-reporting. In the context of this paper, the term symptom validity test (SVT) refers "to any psychometric test, score, or indicator used to detect invalid performance on measures of cognitive or physical capacity, or exaggeration of subjective symptoms" (Greve, Bianchini, & Brewer, 2013). In this sense, the term SVT is used as the superordinate which includes both performance validity tests (PVTs)

\*Correspondence concerning this article should be sent to Peter Giger. Herzogenacker CH-3654 Gunten, 22, Switzerland. Email: p.giger@gmx.net

and self-report validity tests (SRVTs). It reflects the current consensus in literature to use a multi-method approach in forensic evaluations where PVTs and SRVTs are included (Bush et al., 2005; Iverson, 2006). Concerning the incremental validity, the SVTs in a test battery should not correlate highly with one another.

Ideally, SVTs should be sensitive to underperformance or over-reporting but, at the same time, be insensitive to any other factors that may influence test performance (e.g., Hartman, 2002), such as genuine cognitive or mental pathology, age, or education. In fact, a number of studies have shown such insensitivity with selected tests and factors. For instance, SVT usage in childhood age has shown promising results with most stand-alone instruments (e.g., Blaskewitz, Merten, & Kathmann, 2008; Carone, 2008; MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson, 2009); education has no influence on Word Memory Test cutoff-based classification (Rienstra, Spaan, & Schmand, 2009); Test of Memory Malingering performance appears to be independent of older age and education (Tombaugh, 1997) and relatively resistant against mild dementia (Rudman, Oyebode, Jones, & Bentham, 2011), mild mental retardation (Simon, 2007), depression, and pain (Iverson, Le Page, Koehler, Shojania, & Badii, 2007). Other studies showed that age or severe cognitive impairment might, in fact, produce false positive results when the diagnostic decision is based on recommended cutoffs. Therefore, the authors of the Amsterdam Short-Term Memory Test (Schmand & Lindeboom, 2005) recommended not to give the test to patients with clinically obvious symptoms; Reliable Digit Span is well known to be dependent upon age and cognitive impairment (e.g., Blaskewitz et al., 2008; Kiewel, Wisdom, Bradshaw, Pastorek, & Strutt, 2012); performance on the Morel Emotional Numbing Test can be compromised in respondents of older age and with gross cognitive impairment (Morel, 2010).

While there is a growing body of data investigating the influence of genuine cognitive or mental impairment on SVT results, little systematic data is available about the effect of demographic variables on non-patient populations. Disperse data stem mostly from healthy full effort groups in experimental or test validation studies (e.g., Courtney, Dinkins, Allen, & Kuroski, 2003; Rienstra et al., 2009; Teichner & Wagner, 2004). The current study aimed to examine the influence of demographic variables on SVT scores. A selection of PVTs and SRVTs was given to an adult population-based sample of native German speakers of Swiss nationality. A related question to be answered by the study was to investigate rates of positive SVT results in a healthy sample. Positives may be false positives, but they could also signal underperformance or over-reporting of a portion of participants.

Although reference data of this kind may bear little *direct* relevance to clinical and forensic application, they may be important insofar as they provide information about the quality of SVTs. Resulting classifications as to the presence of underperformance or symptom over-reporting should be independent of factors such as gender, age, education, or intelligence and should give information about the validity of test profiles. In other words, SVTs must be constructed in a way that all or almost all healthy test takers score negative on them, *provided that* they employ full test effort in PVTs and respond honestly in SRVTs. Recently, Berthelson, Mulchan, Odland, Miller, and Mittenberg (2013) have highlighted the practical relevance of data from healthy people with normal effort.

## Method

### Participants

A group of 100 volunteer native German speakers from 18 to 60 years of age was studied. In Switzerland, four official languages are spoken. The predominant language spoken by about three quarters of all Swiss citizens is German. To obtain a representative sample of

the Swiss German population, data from the Swiss Federal Statistical Office (Schweizerisches Bundesamt für Statistik, 2011) were consulted. The sample composition aimed to be representative in terms of age, sex, and education.

First, the population was segmented and the proportional number of participants for each segment (defined by age group and education) was computed (Table 1). According to these criteria research participants were recruited by the first author. Distribution of male and female participants was held equal throughout all segments, as far as feasible. The resulting sample consisted of 49 women and 51 men with an average age of 39.4 years ( $SD = 11.9$ ).

Exclusion criteria were: Mental retardation, history of brain injury, history of psychiatric treatment or diagnosis of a mental disorder (except for depression), and alcohol dependence. Moreover, psychologists and undergraduates of psychology were not eligible for inclusion. All participants who agreed to be tested and who did not meet any of the exclusion criteria finished testing, so no data were excluded from the final analyses.

A rough *post hoc* check of verbal intelligence (Vocabulary Test, see Instruments section) resulted in a mean VIQ score of 104.2 ( $SD = 10.6$ ) for the total sample; the distribution conformed to normality.

**Table 1**  
Descriptive Sample Characteristics ( $N = 100$ )

Age group	Level of education					Total
	Primary	Secondary		Tertiary		
	Primary	Vocational	High school	PET/ College	University	
18-20	1	3	2	0	0	6
21-25	1	4	1	1	3	10
26-30	2	4	2	0	3	11
31-35	1	5	1	1	3	11
36-40	2	5	1	1	3	12
41-45	2	6	2	1	3	14
46-50	3	6	2	1	2	14
51-55	2	6	1	1	2	12
56-60	3	5	1	0	1	10
Total	17	44	13	6	20	100

Note. Primary = Elementary/primary school, 9 years of education; Vocational = Vocational school, 12-13 years of education; High school = High school level, 13 years of education; PET = College of professional education and training, 15-16 years of education; University = University level, 17-19 years of education.

### Procedure

Data collection took place between April 2012 and February 2013. All participants were tested individually by the first author. All participants were instructed to answer honestly in questionnaires and to give their best effort in performance tests. To ensure cooperation, every participant received \$20 as an incentive and they were told that the one who showed the best performance would get a voucher of another \$200. Before testing, the purpose of the study was described as gathering reference data for a number of psychological tests, without giving further specification. More detailed feedback about the purpose of the study and the general and the individual results was available on demand after conclusion of the study.

The test battery was always given in the same order of presentation, as described below. Testing lasted between 60 and 90 minutes.

**Instruments**

1) The Medical Symptom Validity Test (MSVT, Green, 2004) Learning Trials and Immediate Recognition (IR). The computerized German test version was used. Ten word pairs are presented on the computer screen, such as “lemon tree” and “hair cut”. IR consists of choosing one of the words in the list (e.g., “lemon”) when given a pair of words containing one list word and one foil word (e.g., “lemon” and “orange”). The recommended cutoff of < 90 was used.

2) The Structured Inventory of Malingered Symptomatology (SIMS, Widows & Smith, 2005; German version, Cima et al., 2003) is a 75-item questionnaire developed to assess endorsement of bizarre, unlikely, or rare symptoms. The items are combined to five subscales tapping the following domains: Low Intelligence, Affective Disorders, Neurological Impairment, Psychosis, and Amnesic Disorders. For use in Switzerland, minor adaptations had to be made for three items as previously described by Giger, Merten, Merckelbach, and Oswald (2010). The total score is the sum of endorsed atypical symptoms. For the total SIMS score we used the cutoff proposed by Cima et al. (2003) for the German version (> 16).

3) The MSVT, Delayed Recognition (DR), Paired Associates (PA), and Free Recall (FR) trials. The recognition task is the same type as described above for IR, but different foil words are used. A consistency score (CNS) from the IR and DR trials is calculated. The cutoff for DR and consistency is < 90. IR is followed by the PA subtest where the person is told the first word and is asked to add the second word (e.g., “lemon”, to which the correct response would be “tree”). In the FR subtest, the person is asked to say as many list words as possible, in pairs or singly. In cases where any one of three symptom validity indicators (IR, DR, and CNS) is positive, a profile analysis is performed in order to distinguish between a profile of genuine cognitive impairment and a profile of underperformance (for details, see Henry, 2009; Howe, Anderson, Kaufman, Sachs, & Loring, 2007).

4) A German-language multiple-choice Vocabulary Test (Wortschatztest, WST, Schmidt & Metzler, 1992) consists of 42 items which are made up of one target word, plus five distractors (pseudo-words). The number of correctly identified words is often used as a rough estimate of verbal intelligence and premorbid cognitive functioning in German-speaking countries. *Post hoc* analysis showed that, in the present sample, verbal intelligence estimates correlated at .55 with education level.

5) The German version of the Amsterdam Short-Term Memory Test (ASTM, Schmand & Lindeboom, 2005) is a PVT in multiple-choice format (three-in-five). For each of the 30 items, five examples of a common category (such as vehicles, colors, trees) are presented on the computer screen. Then, a mathematical problem is presented to increase perceived difficulty. Another five words of the same category are then presented on the screen. The task consists in recognizing which three words are identical to those in the first list. The maximum score is 90 points; the recommended cutoff of < 85 was used in this study. An optional discontinue rule after 15 items was not used in this study.

6) The German adaptation of the Morel Emotional Numbing Test (MENT, Morel, 2010) is a 60-item forced-choice PVT specifically designed for detecting feigned posttraumatic stress disorder. The concept of the test refers to the PTSD symptom of emotional numbing. The task consists in correctly combining emotional expressions on faces with adjectives describing these emotions. Patients with false claims of PTSD may display implausible difficulties in the perception of emotions. The recommended cutoff of > 7 (and > 8 for ages 60 and above) was used in this study.

7) The WAIS-R subtests Digit Span Forward and Backward (Wechsler, 1981). A number of different embedded symptom validity indicators derived from test performance in the Digit subtests of the Wechsler memory or intelligence scales have been proposed. In the current study, we used the Reliable Digit Span (RDS) as first described

by Greiffenstein, Baker, and Gola (1994). Revised cutoffs were later proposed to reduce elevated false-positive rates in different populations. Most authors today appear to recommend a cutoff of < 7, which was also used in the present study (e.g., Greve et al., 2013; Harrison, Rosenblum, & Currie, 2010; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012; Suhr & Barrash, 2007). In patient populations with severe cognitive impairment an even lower cut score should be used. As a second embedded PVT we used the Maximum Span Forward (MSF). As with RDS, different cutoffs were proposed; for the present analyses we used the cut score of MSF < 5 as proposed by Iverson and Tulsky (2003).

8) Finally, a newly developed, but yet unpublished SRVT was included: the List of Indiscriminate Psychopathology (LIPP, Merten & Stevens, 2012). Data collection occurred in the context of ongoing test development, so results will be presented elsewhere.

**Results**

A summary of the results in the six SVTs analyzed in this study is given in Table 2. As was expected, most participants had no difficulties in passing PVTs, so the rates of positive results was low and amounted to a maximum of 4% for the MSF. Only one out of 100 participants scored positive on the SIMS, indicating elevated endorsement of atypical symptoms. For the MSVT, an analysis of primary symptom validity indicators (IR, DR, and CNS) would classify three cases as positives. In the subsequent analysis, one of these cases was categorized as a profile of genuine cognitive impairment. Therefore, in the final MSVT analysis this would be a negative result in terms of possible underperformance.

Nine participants scored positive on only one SVT while three obtained positive scores on two tests (Table 3). In Table 4, the exact frequency distributions are presented for all relevant symptom validity variables.

**Table 2**  
Descriptive Parameters

Test	Variable	Range	M (SD)	Positives
MSVT	IR	90–100	99.4 (1.8)	0
	DR	85–100	99.2 (2.7)	2
	CNS	85–100	98.7 (3.1)	2
	Failure on IR, DR and/or CNS			3
	Profile analysis			2
ASTM		84–90	88.5 (1.5)	3
MENT		0–8	2.1 (1.6)	1
SIMS		0–35	5.2 (4.3)	1
Digit Span	RDS	6–14	9.4 (1.8)	3
	MSF	4–8	6.2 (0.9)	4

Note. MSVT = Medical Symptom Validity Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; ASTM = Amsterdam Short-Term Memory Test; MENT = Morel Emotional Numbing Test; SIMS = Structured Inventory of Malingered Symptomatology; RDS = Reliable Digit Span; MSF = Maximum Span Forward.

**Table 3**  
Number of positive classifications

Number of positive tests	0	1	2	3	4	5	6
n	88	9	3	0	0	0	0

**Table 4**  
Frequency Distribution in the Total Sample

Raw scores	MSVT			ASTM		MENT		SIMS		Digit Span			
	IR	DR	CNS	Raw scores	Errors	Total score	RDS	MSF					
100	88	90	81	90	29	0	15	0	2	15	0	8	10
95	11	6	14	89	30	1	27	1	8	14	3	7	20
90	1	2	3	88	21	2	20	2	12	13	3	6	51
85	0	2	2	87	10	3	22	3	17	12	8	5	15
80	0	0	0	86	6	4	11	4	10	11	12	4	4
75	0	0	0	85	1	5	2	5	17	10	14	3	0
70	0	0	0	84	3	6	2	6	9	9	27	2	0
65	0	0	0	83	0	7	0	7	6	8	19	1	0
60	0	0	0	82	0	8	1	8	4	7	11	0	0
< 60	0	0	0	< 82	0	> 8	0	9	4	6	3		
								10	4	5	0		
								11	3	4	0		
								12	1	3	0		
								13	0	2	0		
								14	2	1	0		
								> 14	1 <sup>1</sup>	0	0		

Note. <sup>1</sup>One participant scored 35 on the SIMS. MSVT = Medical Symptom Validity Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; ASTM = Amsterdam Short-Term Memory Test; MENT = Morel Emotional Numbing Test; SIMS = Structured Inventory of Malingered Symptomatology; RDS = Reliable Digit Span; MSF = Maximum Span Forward.

To investigate the effects of demographic variables, multiple linear regression analyses were performed with SVT indicators as dependent variables. Besides age, education, and sex, we decided to include verbal intelligence estimates (as measured by the Vocabulary Test, WST) as independent variables. Sex and education had no effect on SVT scores (Table 5). Age was a relevant factor for MSVT IR and CNS scores, and for MENT score, with a tendency to lower performance at older age. Higher verbal intelligence was associated with higher MSVT IR and CNS scores, higher ASTM scores with fewer errors in MENT, and decreased symptom endorsement in SIMS with higher RDS and MSF scores.

**Table 6**  
Test Scores (Mean and Standard Deviations) for Five Symptom Validity Tests

Level of education	Age group	ASTM	MENT	SIMS	RDS	MSF
Primary	18–40	87.3 (1.4)	2.3 (2.0)	7.9 (4.1)	8.9 (1.9)	5.7 (1.1)
	41–60	87.5 (1.5)	3.0 (2.3)	9.5 (9.5)	9.6 (2.0)	6.6 (0.8)
Secondary	18–40	88.5 (1.7)	1.3 (1.3)	5.6 (3.0)	9.2 (1.6)	6.2 (0.6)
	41–60	88.8 (1.3)	2.5 (1.5)	4.4 (2.8)	9.2 (1.8)	6.0 (1.1)
Tertiary	18–40	88.7 (1.2)	1.9 (1.1)	3.6 (2.1)	9.5 (1.8)	6.1 (1.1)
	41–60	89.3 (0.5)	2.4 (1.1)	2.9 (1.4)	10.4 (2.3)	6.5 (1.0)

Note. ASTM = Amsterdam Short-Term Memory Test; MENT = Morel Emotional Numbing Test; SIMS = Structured Inventory of Malingered Symptomatology; RDS = Reliable Digit Span; MSF = Maximum Span Forward.

**Table 5**  
Test Scores (Means and Standard Deviations) for the Subtests of the Medical Symptom Validity Test Stratified by Education and Age Groups

Level of education	Age Group	IR	DR	CNS	PA	FR
Primary	18–40	99.3 (1.9)	97.9 (5.7)	98.6 (3.8)	78.6 (23.4)	65.0 (10.0)
	41–60	98.0 (3.5)	98.0 (4.8)	96.5 (5.3)	74.0 (24.1)	69.0 (13.5)
Secondary	18–40	99.8 (0.9)	99.6 (1.3)	99.5 (1.6)	93.6 (11.9)	82.3 (11.2)
	41–60	99.1 (1.9)	99.5 (2.0)	98.6 (3.2)	90.7 (12.8)	77.6 (14.2)
Tertiary	18–40	99.7 (1.3)	99.0 (2.8)	98.7 (3.0)	96.7 (8.2)	84.0 (12.6)
	41–60	99.6 (1.5)	99.6 (1.5)	99.1 (2.0)	97.3 (6.5)	81.8 (17.4)

Note. IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; PA = Paired Associates; FR = Free Recall.

Tables 6 and 7 present means and standard deviations of the SVTs, stratified by age (dichotomized into younger and older age group) and education (grouped into primary, secondary, and tertiary level). Table 8 shows the correlations between the SVT measures. Most correlations between independent SVTs were low, often not reaching the threshold of statistical significance.

## Discussion

The present study aimed to collect reference data for six SVTs on a population-based representative sample of adult participants of working age. To our knowledge, no such representative data are yet available for any of the tests used in this study. The

**Table 7**  
Linear Regression Analysis: Contribution of Sex, Age, Education, and Verbal Intelligence to Test Scores

SVT	Dependent variable	Independent variable <sup>1</sup>	B	t	R <sup>2</sup>
MSVT	IR	Constant	95.50	48.06*	.12
		Sex	-0.34	-0.95	
		Age	-0.04	-2.24*	
		Education	-0.08	-0.47	
		Verbal intelligence	0.06	2.71*	
MSVT	DR	Constant	97.08	31.59*	.05
		Sex	-0.78	-1.42	
		Age	-0.02	-0.70	
		Education	0.02	0.09	
		Verbal intelligence	0.04	1.15	
MSVT	CNS	Constant	93.96	28.01*	.10
		Sex	-0.86	-1.43	
		Age	-0.06	-2.07*	
		Education	-0.14	-0.50	
		Verbal intelligence	0.08	2.32*	
ASTM	Total score	Constant	83.55	55.16*	.19
		Sex	0.15	0.53	
		Age	0.02	1.20	
		Education	0.24	1.92	
		Verbal intelligence	0.03	2.11*	
MENT	Error score	Constant	5.52	3.34*	.15
		Sex	-0.12	-0.40	
		Age	0.04	3.33*	
		Education	-0.21	1.55	
		Verbal intelligence	-0.05	-3.04*	
SIMS	Total score	Constant	16.14	3.60*	.18
		Sex	0.61	0.76	
		Age	0.02	0.62	
		Education	-0.63	-1.73	
		Verbal intelligence	-0.11	-2.23*	
RDS	Total score	Constant	2.05	1.07	.17
		Sex	-0.08	-0.23	
		Age	-0.01	-0.72	
		Education	-0.06	-0.40	
		Verbal intelligence	0.08	3.79*	
MSF	Total score	Constant	4.21	4.00*	.07
		Sex	-0.18	-0.97	
		Age	-0.01	-0.57	
		Education	-0.05	0.60	
		Verbal intelligence	0.03	2.20*	

Note. <sup>1</sup>Sex was coded as 1 = female and 2 = male; education was coded from 1 (primary education) to 5 (university degree). MSVT = Medical Symptom Validity Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; ASTM = Amsterdam Short-Term Memory Test; MENT = Morel Emotional Numbing Test; SIMS = Structured Inventory of Malingered Symptomatology; RDS = Reliable Digit Span; MSF = Maximum Span Forward.

\* $p < .05$ .

rationale of the selection was 1) to cover a spectrum of instruments that tap both underperformance and symptom over-reporting, 2) to resort to both stand-alone and embedded

measures, and 3) to select instruments that are in practical use in German speaking countries.

Recently, Rienstra et al. (2009) have published reference data for the Dutch version of the Word Memory Test (Green, 2003). As was demonstrated by these authors, and was previously shown by Schmand and Lindeboom (2005) for the ASTM, symptom validity tests appear to be resilient against distorting effects when adapted into other languages and cultures (at least within the Western world and provided the adaptations were thoroughly performed). The primary reason for this may be the fact that traditional stand-alone PVTs have such a low test ceiling that the test performance is, to a considerable degree, resistant against many potential influencing factors (including the effects of brain damage).

Some factors that potentially influence both test performance and self-report measures are demographic variables such as sex, age, education, and person variables such as intelligence. Another purpose of this study was to investigate the effects of these factors on SVT scores. In fact, significant effects on a number of SVTs were found for both age and a rough measure of verbal intelligence. For the Test of Memory Malingering (TOMM), Tombaugh (1997) described effects of age and education in a non-patient population. However, a systematic body of data appears to be missing because only few studies focus on SVT performance in fully cooperative, healthy people. Their performance is usually only of interest when compared to other groups such as experimental malingerers. Although Tombaugh found that age and education of healthy respondents accounted for some variance in the TOMM, classification accuracy was not impacted by these factors. This was also true for the present study: despite the significant effects of age and/or verbal intelligence on a number of SVT scores, the recommended cutoff-based decision making appears to be widely resistant against these effects in that there was no elevated rate of positive SVT scores in our sample.

In contrast to Tombaugh (1997), we included a measure of verbal intelligence, the effect of which corrected the effect of education to an extent that the latter was not detectable anymore. In fact, preliminary analyses in our data had shown some effect of education on selected SVTs when regression analyses were performed on age, sex, and education alone.

Low positive rates were found for all SVTs employed. The maximum rate amounted to 4% for the MSF. It is impossible to determine, with a sufficient degree of certainty, if the positives in this study were false positives or true positives. We performed an inspection of the 12 cases that scored positive on at least one SVT. With one case, the individual test profile as well as extreme elevations both on the SIMS and on pseudosymptom scales of the newly developed List of Indiscriminate Psychopathology indicated a high probability of symptom over-reporting. Two more test profiles were judged to be more likely to reflect negative response bias than valid results. This would result in a rate of true positive classifications of invalid test profiles for three out of 100 cases of the reference sample. For the remaining nine cases, close inspection showed that they were more likely to represent false positives than true positives. However, this *post hoc* classification must be considered as hypothetical. In contrast to real-world clinical and forensic contexts, no plausibility and consistency check could be performed that would have encompassed other sources (such as comprehensive neuropsychological test data, medical history, records, behavioral observations, reliable third party information).

With a supposed rate of 3% of participants who truly underperformed and/or over-reported at least in part of the test battery, numbers of hypothetical false positives can be derived from Table 9 based on adjusted numbers of Table 2.

With any normative or reference sample, it cannot be expected that all participants respond honestly in self-report measures and work with full effort in performance tests. The rate of

**Table 8**  
Pearson Product-Moment Correlations between Symptom Validity Tests

		MSVT	MSVT	MSVT	ASTM	MENT	SIMS	RDS	MSF
		IR	DR	CNS	Total score	Error score	Total score	Total score	Total score
MSVT	IR	1	.15	.61*	.14	-.45*	-.06	-.03	-.02
MSVT	DR	.15	1	.81*	.19	-.21*	-.41*	.20	.09
MSVT	CNS	.61*	.81*	1	.20	-.35*	-.35*	.13	.08
ASTM	Total score	.14	.19	.20	1	-.22*	-.24*	.32*	.26*
MENT	Error score	-.45*	-.21*	-.35*	-.22*	1	.14	-.01	-.00
SIMS	Total score	-.06	-.41*	-.35*	-.24*	.14	1	-.13	-.03
RDS	Total score	-.03	.20	.13	.32*	-.01	-.13	1	.64*
MSF	Total score	-.02	.09	.08	.26*	-.00	-.03	.64*	1

Note. MSVT = Medical Symptom Validity Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; ASTM = Amsterdam Short-Term Memory Test; MENT = Morel Emotional Numbing Test; SIMS = Structured Inventory of Malingered Symptomatology; RDS = Reliable Digit Span; MSF = Maximum Span Forward.  
\* $p < .05$ .

**Table 9**  
Hypothetical False-Positives in Symptom Validity Tests

Test	Variable	Positives	Hypothetical false-positive rate	Number of hypothetically false-positive test results	<i>n</i>
MSVT	IR	0	0	0	91
	DR	2	0	1	8
	CNS	2	1	2	1
	Failure on IR, DR and/or CNS	3	1	> 2	0
	Profile analysis	2	0		
ASTM		3	2		
MENT		1	1		
SIMS		1	0		
Digit Span	RDS	3	2		
	MSF	4	4		

Note. MSVT = Medical Symptom Validity Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; ASTM = Amsterdam Short-Term Memory Test; MENT = Morel Emotional Numbing Test; SIMS = Structured Inventory of Malingered Symptomatology; RDS = Reliable Digit Span; MSF = Maximum Span Forward.

underperformance and over-reporting in such samples varies depending on diverse factors, such as positive incentives to fully cooperate, restricted voluntariness of participation, qualification of the experimenter, perceived fairness of the experimenter's behavior toward participants, perceived relevance of the study, or expected implicit or explicit consequences of test results. In undergraduate populations, a higher percentage of restricted cooperation must be expected if students are obliged to participate in studies in order to gain credits (An, Zakzanis, & Joordens, 2012).

As a consequence, normative and reference data are contaminated by invalid data to an unknown degree. Even if SVTs were included in such studies to adjust for invalid data, this would not solve the problem of deciding if positive results are exclusively false positives or true positives, or which cases are false and which are true positives. Rienstra et al. (2009) decided to exclude two participants who scored positive on the WMT validity indicators from their reference data, whereas we decided not to exclude any case. There are positive arguments for both procedures. Both will potentially lead to distortions in the accuracy of the resulting reference data as long as there is no definite criterion to distinguish between false and true positives.

Limitations of the study arise primarily from the restricted number of participants. As a result, the data should not be read as normative, but as what they are: Reference data. The reliability of our estimates of the influence of demographic variables on SVT scores is limited by reduced variance of the latter in healthy participants. The low variance of SVT scores also attenuates the

intercorrelations between SVTs. The question of generalizability of Swiss reference data to other countries of German language and, more importantly, to non-German countries is obvious although, as mentioned in the introduction, there is some reason to suppose that SVT research is relatively robust against language and cultural distortions within the Western civilization.

### Conflicts of interest

None for the first author. The second author has participated in the German adaptation of numerous symptom validity tests, among those mentioned in the paper: the MENT, the ASTM, the MSVT, and WMT. Royalties are received only for sales of the German ASTM version.

### Notes

<sup>1</sup>The paper was first presented as a poster and data blitz presentation at the Third European Symposium on Symptom Validity Assessment in Wuerzburg, Germany, in June 2013.

### References

- An, K. Y., Zakzanis, K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, 27, 849-857.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, 27, 909-916.

- Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology*, 23, 379-391.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ... Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, 20, 419-426.
- Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury*, 22, 960-971.
- Carone, D. A., Bush, S. S., & Iverson, G. L. (2013). Providing feedback on symptom validity, mental health, and treatment in mild traumatic brain injury. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury; Symptom validity assessment and malingering* (pp. 101-118). New York: Springer.
- Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B., ... Merkelbach, H. (2003). Strukturierter Fragebogen Simulierter Symptome – Die deutsche Version des Structured Inventory of Malingered Symptomatology: SIMS [German version of the SIMS]. *Der Nervenarzt*, 74, 977-986.
- Courtney, J. C., Dinkins, J. P., Allen, L. M., & Kuroski, K. (2003). Age related effects in children taking the Computerized Assessment of Response Bias and World Memory Test. *Child Neuropsychology*, 9, 109-116.
- Dandachi-Fitzgerald, B., & Merkelbach, H. (2013). Feigning ≠ feigning a memory deficit: The Medical Symptom Validity Test as an example. *Journal of Experimental Psychopathology*, 4, 46-63.
- Giger, P., Merten, T., Merkelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice*, 10, 440-463.
- Göbber, J., Petermann, F., Piegza, M., & Kobelt, A. (2012). Beschwerdvalidierung bei Rehabilitanten mit Migrationshintergrund in der Psychosomatik [Symptom validation in patients with migration background in psychosomatic medicine]. *Die Rehabilitation*, 51, 356-364.
- 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: Green's Publishing.
- Greiffenstein, M., Baker W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6, 218-224.
- Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *The Clinical Neuropsychologist*, 27, 108-137.
- Harrison, A. G., Rosenblum, Y., & Curry, S. (2010). Examining unusual Digit Span performance in a population of postsecondary students assessed for academic difficulties. *Assessment*, 17, 283-293.
- Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing. Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology*, 17, 709-714.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093-1129.
- Henry, M. (2009). Beschwerdvalidierungstests in der zivil- und sozialrechtlichen Begutachtung: Verfahrensüberblick [Symptom validity tests in civil forensic assessment: A review of instruments]. In T. Merten & H. Dettenborn (Eds.), *Diagnostik der Beschwerdvalidität* (pp. 118-161). Berlin, Germany: Deutscher Psychologen Verlag.
- Howe, L. L. S., Anderson, A. M., Kaufmann, 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.
- Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology*, 13, 77-90.
- Iverson, G. L., & Tulskey, D. S. (2003). Detecting malingering on the WAIS-III unusual Digit Span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology*, 18, 1-9.
- Iverson, G. L., Le Page, J., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist*, 21, 532-546.
- Kiewel, N. A., Wisdom, N. M., Bradshaw, M. R., Pastorek, N. J., & Strutt, A. M. (2012). A retrospective review of Digit Span-related effort indicators in probable Alzheimer's disease patients. *The Clinical Neuropsychologist*, 26, 965-974.
- MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology*, 16, 521-531.
- Merten, T., & Stevens, A. (2012). *List of Indiscriminate Psychopathology (LIPP)*. Unpublished manuscript.
- Morel, K. R. (2010). *MENT – Manual for the Morel Emotional Numbing Test for posttraumatic stress disorder (2nd edition)*. Author.
- Rienstra, A., Spaan, P. E. J., & Schmand, B. (2009). Reference data for the World Memory Test. *Archives of Clinical Neuropsychology*, 24, 255-262.
- Rudman, N., Oyeboode, J. R., Jones, C. A., & Bentham, P. (2011). An investigation into the validity of effort tests in a working age dementia population. *Aging & Mental Health*, 15, 47-57.
- Schmand, B., & Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test. Manual*. Leiden, NL: PITS.
- Schmidt, K.-H. & Metzler, P. (1992). *Wortschatztest (WST) [Vocabulary Test]*. Weinheim: Beltz.
- Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable Digit Span: A systematic review and cross-validation study. *Assessment*, 19, 21-30.
- Schweizerisches Bundesamt für Statistik (2011). *Statistisches Jahrbuch der Schweiz 2011 [Statistical yearbook 2011]*. Zürich: Neue Zürcher Zeitung.
- Simon, M. J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of Clinical Psychology*, 63, 339-344.
- Suhr, J. A., & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In G. L. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 131-179). Oxford: University Press.
- Sweet, J. J., & Guidotti-Bretting, L. M. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology*, 4, 6-19.
- 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.
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9, 260-268.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised manual*. New York: The Psychological Corporation.
- Widows, M. R., & Smith, G. P. (2005). *SIMS-Structured Inventory of Malingered Symptomatology. Professional Manual*. Lutz, FL: Psychological Assessment Resources.