

The Self-Report Symptom Inventory (SRSI): a New Instrument for the Assessment of Distorted Symptom Endorsement

Thomas Merten¹ · Harald Merckelbach² · Peter Giger³ · Andreas Stevens⁴

Received: 4 January 2016 / Accepted: 24 April 2016 / Published online: 12 May 2016
© Springer Science+Business Media New York 2016

Abstract Self-report instruments to detect distorted symptom reporting play a crucial role in clinical and forensic psychology. Most of the instruments currently available for this purpose only list implausible symptoms, which makes them easily identifiable as symptom validity tests. We developed the Self-Report Symptom Inventory (SRSI), combining five self-report scales of genuine symptoms with five pseudosymptom scales to screen for distorted symptom reporting in various domains (e.g., depression, post-traumatic stress). With a preliminary questionnaire version, we collected data in a heterogeneous sample ($N=239$) and performed an item selection, resulting in the final 107-item version. This version was evaluated in civil forensic patients, inmates of a prison, and a population-based sample; ($N=387$). Data show that (a) SRSI pseudosymptom scores correlate highly ($\geq .80$) with other instruments tapping distorted symptom endorsement, notably the Structured Inventory of Malingered Symptomatology; (b) High SRSI pseudo-

symptom scores tend to correlate with underperformance; and (c) The psychometric features of the SRSI are satisfactory, with internal consistency for the total scales $>.90$ and retest reliability $>.85$. The instrument appears to be a promising tool for examining symptom exaggeration, but further work is required, in particular cross-validation with other samples and different methods.

Keywords Psychological assessment · Questionnaire · Symptom validity testing · Malingering · Symptom overreporting · Forensic assessment · Self-Report Symptom Inventory (SRSI)

Testing the credibility of reported symptoms is a core issue in forensic evaluations (Bush et al., 2005; Bush, Heilbronner, & Ruff, 2014; Heilbronner, Sweet, Morgan, Larrabee, Millis, & Conference Participants, 2009; see also Young, 2014, for an ethical perspective); growing interest is also evident in clinical and rehabilitation contexts (e.g., Carone, Bush, & Iverson, 2013; Göbber, Petermann, Piegza, & Kobelt, 2012). To evaluate the credibility of symptoms, experts may rely on two types of instruments, namely performance validity tests (PVTs) that aim to detect underperformance and self-report validity tests (SRVTs) that evaluate possible indiscriminate symptom endorsement. In the extant literature, both types of tools are traditionally referred to as symptom validity tests or SVTs (although the recent literature partly restricts this term to SRVTs, cf. Larrabee, 2012).

SRVTs may be informative in forensic contexts where secondary gain motives (e.g., compensation money, reduced criminal responsibility, stimulant medication) are often present. Indeed, many studies have documented that distorted symptom presentation occurs on a non-trivial scale in civil and criminal forensic evaluations, with prevalence rates often

Electronic supplementary material The online version of this article (doi:10.1007/s12207-016-9257-3) contains supplementary material, which is available to authorized users.

✉ Thomas Merten
thomas.merten@vivantes.de

- ¹ Department of Neurology, Vivantes Klinikum im Friedrichshain, Landsberger Allee 49, D-10249 Berlin, Germany
- ² Forensic Psychology Section, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
- ³ Federal Department of Defence, Civil Protection and Sports, Bern, Switzerland
- ⁴ Medizinisches Begutachtungsinstitut, Tübingen, Germany

exceeding 30 %, sometimes even exceeding 50 % (e.g., Ardolf, Denney, & Houston, 2007; Chafetz, Abrahams, & Kohlmaier, 2007; Merten, Thies, Schneider, & Stevens, 2009; Schmand, Lindeboom, Schagen, Heijt, Koene, & Hamburger, 1998). A summary discussion of different base rate estimates in a variety of neuropsychological samples can be found in Boone (2013). Arguably, in legal settings, it is risky to take the authenticity of self-reported symptoms for granted and clinicians are well advised to directly address the issue of symptom validity by testing for response distortions.

A variety of different measurement approaches has been developed to examine distorted symptom presentation. PVTs include forced-choice tests and embedded measures. Adapting the forced-choice format, Morel (2010) developed a test to identify feigned post-traumatic stress disorder. The most important interview-based scales currently available are the Structured Interview of Reported Symptoms (SIRS-2; Rogers, Sewell, & Gillard, 2010) and the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). Self-report measures, on which the following more detailed review will focus, comprise both screens and multiscale inventories.

SRVTs tap into overreporting of symptoms, but they usually do not cover distortions such as random responding and response inconsistency nor do they measure social desirability, fake good tendencies, and other forms of defensiveness. Several SRVTs have been developed to assess negative response bias in symptom reporting (see Smith, 2008; Young, 2014, for reviews). One important category comprises scales that are embedded into more comprehensive personality and clinical inventories such as those of the Minnesota Multiphasic Personality Inventory (MMPI; e.g., Arbisi & Ben-Porath, 1995; Wiggins, Wygant, Hoelzle, & Gervais, 2012). The MMPI-2 and MMPI-2-RF validity scales have repeatedly been shown to be sensitive to symptom overreporting and feigning (e.g., Bolinger, Reese, Suhr, & Larrabee, 2014; Rogers, Sewell, Martin, & Vitacco, 2003). The Personality Assessment Inventory (PAI; Morey, 2007) and its adolescence version also have embedded scales for the detection of overreporting. A number of studies found that the PAI validity scales and indices are useful (e.g., Rios & Morey, 2013; Rogers, Gillard, Wooley, & Ross, 2012; Vossler-Thies, Stevens, Engel, & Licha, 2013).

Another category of SRVTs comprises the stand-alone questionnaires specially designed for detecting distorted symptom report. According to surveys by Dandachi-FitzGerald, Ponds, and Merten (2013) and Martin, Schroeder, and Odland (2015), the most widely used stand-alone SRVT among European and North American neuropsychologists is the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005). Its five subscales consist of 75 items addressing bizarre, rare, atypical,

or extreme symptoms ostensibly fitting into broad pathological domains (Psychosis, Neurological Impairment, Amnesic Disorders, Low Intelligence, and Affective Disorders). The subscale scores are summed up to obtain a total score indicating possible negative response bias. In a recent meta-analytic review, van Impelen, Merckelbach, Jelacic, and Merten (2014) assembled 31 published empirical studies with data from 49 subsamples and 4869 SIMS protocols. The authors concluded that the SIMS has several strong features (e.g., high sensitivity to distorted symptom reporting), but they also identified important limitations of the instrument. One of them appears to be a limited specificity in samples with genuine severe psychopathology. Thus, genuine patients may be wrongly classified as feigning when the test user relies on conventional cut-offs. Another weakness of the SIMS is its exclusive reliance on bizarre or atypical symptoms, which may make the instrument easily recognizable as an SRVT. This problem is also evident for a recently developed short version of the SIMS (Malcore, Schutte, Van Dyke, & Axelrod, 2015).

Below, we describe the construction and validation of a new instrument for the identification of distorted symptom report. With a number of studies going on in different European countries, the validation of this instrument is work in progress. Given the complexity of both the development of psychometrically sound instruments and the different approaches to determine classification accuracy of SVTs (cf. Nies & Sweet, 1994; Rogers, 2008; Vitacco, Rogers, Gabel, & Munizza, 2007), the following analyses are mostly limited to the testing of the new instrument against an external criterion, the SIMS.

Rationale and Questionnaire Construction

Given the paucity of stand-alone symptom overreporting instruments that combine common symptoms with bogus symptoms and limited data available on most self-report validity instruments in some major European languages, we developed the Self-Report Symptom Inventory (SRSI). The SRSI was first conceived in 2006 after performing an empirical analysis of the SIMS and identifying weaknesses of the instrument both at a general level and at the level of individual items (Merten, Friedel, & Stevens, 2007; van Impelen et al., 2014; see also Santamaría Fernández, 2014, for a comprehensive analysis). It combines genuine clinical and pseudosymptom (i.e., validity) scales. Honest patients with genuine symptomatology are expected to endorse clinical symptoms but not pseudosymptoms. Similar to the SIMS items, the SRSI pseudosymptoms were selected by experts as presenting bizarre, atypical, or rare complaints that, however, in the eyes of laypersons, appear to belong to known syndromes. At the same time, the items should stand the test of an empirical item analysis, with an explicit external selection criterion.

The rationale for the inclusion of genuine symptom scales was threefold: First, one of the major flaws of the SIMS (which is the fact that it is easily identified as a symptom validity measure) is camouflaged by the combination of genuine and bogus symptoms in one instrument. Second, with comparative data on the symptom scales in different populations (healthy people, different patient populations), the clinician will be enabled to gather information on individual symptom endorsement along with information about the credibility of symptom report. Third, an index of the ratio between genuine symptom endorsement and number of endorsed bogus item may be developed as an additional parameter of distorted response styles.

In general, SVTs should be sensitive to poor effort or distorted symptom report but, at the same time, they should be relatively insensitive to factors like gender, age, and educational background. Most importantly, patients with genuine complaints who respond honestly should endorse pseudosymptom items only at a low frequency.

The new instruments was to tap a symptom spectrum typically endorsed in the fields of civil litigation, public and administrative law, workers' compensation, and other branches of social security. In those referral contexts, SIMS subscales Psychosis and Low Intelligence do not seem to perform well (Merten et al., 2007) because they tap into blatant, less plausible forms of distorted symptom reporting.

Development of a Preliminary Version

Detailed information about the rational scale construction, the development and the empirical analysis of a preliminary questionnaire version can be found in the [supplemental file](#). For the preliminary SRSI version, we selected 15 items for each genuine symptom scale and for each pseudosymptom scale from the initial item pool. Scores on clinical and bogus symptoms subscales are summed up separately to obtain a total genuine symptom and a total pseudosymptom score, respectively. Furthermore, the SRSI contains two warming-up items checking the a priori affirmation of cooperativeness (item #1: *I have read the instructions above*; item #2: *I am prepared to answer all questions honestly*) and a short 5-item consistency check to gauge careless responding (see Meade & Craig, 2012; Meyer, Faust, Faust, Baker, & Cook, 2013), especially an undifferentiated affirmative tendency (i.e., yeah-saying). Table 1 shows the scale structure of the SRSI.

The preliminary version was tested in a sample of $N=239$ comprising different subsamples of instructed simulators, referrals for independent medical examinations, and neurological patients (*cf.* [supplemental file](#), for more details). For all participants, scores of the German version of the SIMS (Cima et al., 2003) were available. The empirically established cutoff for the German SIMS version is 16, with scores >16 indicating

an elevated probability of feigned symptom report. Item analysis and item selection of the SRSI were based on the results of the participants in the SIMS. On the basis of item-wise correlations with SIMS scores, an item-selection procedure was performed reducing the number of items per subscale from 15 to 10. The final SRSI version comprised 107 items.

The choice of the SIMS for item selection was guided by the fact that there is no other instrument available in German language that is both conceptually very similar to the pseudosymptom scales of the SRSI and in wide use in Germany (*cf.* Dandachi-FitzGerald, Merten, Ponds, & Niemann, 2015). Moreover, the utility of the SIMS as a screener for feigned symptoms in clinical contexts and in independent medical and psychological examinations has been demonstrated in a number of German studies (e.g., Göbber et al., 2012; Zimmermann, Kowalski, Alliger-Horn, Danker-Hopfe, Engers, Meermann, & Hellweg, 2013).

Information on the results with the preliminary version can be found in the [supplemental file](#). Means and standard deviations for the recomputed scale scores (after item reduction) are summarized in Table [Suppl-2](#), grouped by SIMS normal responders vs. SIMS high scorers. On average, the participants of the pooled sample endorsed 32.33 genuine symptoms (range from 3 to 50; $SD=9.84$) and 11.23 pseudosymptoms (range from 0 to 48; $SD=9.54$). The number of endorsed SRSI pseudosymptoms correlated at .81 with SIMS scores.

German Studies with the Final SRSI Version

SRSI in a Population-Based Sample

Sample and Method The third author (PG) collected population-based reference data in a sample of 100 native German speakers of Swiss nationality (Giger & Merten, 2013). Based on demographic data from the Swiss Federal Statistical Office (Schweizerisches Bundesamt für Statistik, 2011), the sample of adults was drawn in a way to be representative in terms of age, gender, and education. Individuals with a known history of brain injury, intellectual disability, prior psychiatric treatment, and alcohol dependence were excluded from participation. The final sample consisted of 51 men and 49 women, with a mean age of 39.43 years (ranging from 18 to 60 years; $SD=11.93$). The sampling method resulted in a group with a rough estimate of verbal intelligence of 104.22 ($SD=10.63$).

All participants were tested individually. The SRSI was given at the end of a battery of tests and questionnaires, mostly tapping symptom validity. The participants were instructed to answer honestly and to perform to the best of their abilities.

Results The majority of the participants (71.0 %) had a total SRSI pseudosymptom score of zero, endorsing none of the

Table 1 Scale structure of the Self-Report Symptom Inventory

Domain	Scale		No. of items	
	No./abbrev.	Name	Preliminary version	Final version
Genuine symptom scales	1 XCo	Cognitive	15	10
	2 XDe	Depressive	15	10
	3 XPa	Pain	15	10
	4 XSo	Nonspecific somatic	15	10
	5 XAn	PTSD/anxiety	15	10
Subtotal	Xx	Total genuine symptoms	75	50
Pseudosymptom scales	6 YCo	Cognitive/memory	15	10
	7 YMo	Neurological: motor	15	10
	8 YSe	Neurological: sensory	15	10
	9 YPa	Pain	15	10
	10 YAn	Anxiety/depression (incl. PTSD)	15	10
Subtotal	Yy	Total pseudosymptoms	75	50
Additional items	A priori cooperativeness	2	2	
	Consistency check	5	5	

presumed atypical or bizarre symptoms. The large majority (98.0 %) endorsed less than 5 pseudosymptoms; 99.0 % of the participants scored 10 or less points on the pseudosymptom scale. A post hoc profile analysis of the one single person in this sample who endorsed 17 pseudosymptoms revealed that he was not a false-positive but a true positive case. This participant also scored 35 points on the SIMS, signaling a high probability of symptom over-endorsement.

Reported symptom levels were also rather low. On average, participants endorsed 7.21 of the 50 symptoms of the SRSI genuine symptom scales ($SD=7.12$). Conversely, endorsement of the consistency check items was relatively high, with an average of 3.8 endorsed items (of the total of five).

Means, standard deviations, and ranges for the individual scale scores are given in Table 2. Contrary to pseudosymptoms, the genuine symptom scales showed more variation in the normal sample. However, as a trend, mean symptom scores in the population-based sample were also low.

Neither total SRSI genuine symptom scores, nor SRSI pseudosymptom scores correlated significantly with age ($r=.07$ and $.03$, respectively). There was a moderate, albeit significant correlation with education, as measured on a five-point academic achievement scale, such that less educated respondents endorsed both more genuine symptoms and more pseudosymptoms on the SRSI ($r=-.29$ and $r=-.24$, respectively). One-way ANOVAs revealed no significant effects of gender on symptom or pseudosymptom endorsement ($F(1, 100)=0.77$ and $F(1, 100)=0.33$, respectively). The correlation between SIMS and total SRSI pseudosymptoms scores was $.80$.

SRSI in Independent Medical Examinations

Sample and Method The SRSI was administered to 207 consecutive patients, with a mean age of 45.46 years ($SD=11.79$; range 18 to 69), who were seen at the private practice of the fourth author (AS). The data were collected between April 2012 and April 2013. The sample was completely distinct from the one used for item analysis and selection (*cf. supplemental file*). The 142 men and 65 women were given a test battery in the context of an independent medical examination. Following the International Standard Classification of Education or ISCED-1997 (United Nations Educational, Scientific and Cultural Organization, 1997) and its specification for Germany (Schroedter, Lechert, & Lüttinger, 2006), the educational background of the participants was as follows: primary ($n=3$), lower secondary ($n=11$), upper secondary ($n=3$), upper secondary, vocational ($n=111$), post-secondary, non-tertiary ($n=14$), first stage of tertiary ($n=34$), first stage of tertiary, university ($n=29$), second stage of tertiary, and doctorate ($n=2$). Referrals were from the German Workers' Compensation Board (47.3 %), private insurance companies (41.0 %), and state agencies (11.1 %). Only cases with complete data sets on the SRSI and SIMS were included in the current sample. A large number of additional test data were available. Results of a performance validity test, namely the Word Memory Test (WMT; Green, 2003), were available for all but two cases. The WMT is a well-researched verbal memory measure that indexes underperformance. It is a forced-choice test with three primary validity indicators (immediate recognition IR, delayed recognition DR, consistency CNS).

Table 2 Means and standard deviations for the SRSI scales scores (population-based normal sample and patient sample from independent medical examinations)

Scale scores (final SRSI version)	Population-based normal sample ($n = 100$)		Patients of independent medical examination sample				Cohens's d	t
			SIMS Normal responders ($n = 138$)		SIMS High scorers ($n = 69$)			
	M (SD)	Range	M (SD)	Range	M (SD)	Range		
Genuine symptom scale total	7.21 (7.12)	0–42	20.97 (10.18)	1–43	37.45 (6.14)	25–49	1.96	14.49*
Cognitive	1.38 (1.75)	0–7	4.10 (3.37)	0–10	8.61 (1.56)	3–10	1.72	13.23*
Depressive	0.86 (1.42)	0–8	3.05 (2.46)	0–9	6.10 (2.17)	2–10	1.31	8.75*
Pain	1.09 (1.60)	0–8	3.51 (3.05)	0–10	6.59 (3.04)	0–10	1.01	6.86*
Nonspecific somatic	1.72 (2.23)	0–10	5.67 (3.30)	0–10	9.22 (1.25)	4–10	1.42	11.14*
PTSD/anxiety	2.16 (2.16)	0–10	4.62 (2.90)	0–10	6.93 (2.16)	2–10	0.90	6.48*
Pseudosymptom scale total	1.99 (0.73)	0–17	2.88 (3.30)	0–19	13.00 (7.11)	0–33	1.83	11.21*
Cognitive	0.18 (0.54)	0–3	0.58 (0.99)	0–5	3.83 (2.30)	0–9	1.84	11.22*
Neurological: motor	0.06 (0.28)	0–2	0.25 (0.54)	0–2	1.72 (1.64)	0–6	1.20	7.24*
Neurological: sensory	0.28 (0.82)	0–5	0.63 (1.02)	0–5	2.39 (2.05)	0–8	1.09	6.75*
Pain	0.08 (0.31)	0–2	0.41 (0.84)	0–4	1.90 (1.99)	0–8	0.98	6.00*
Anxiety/depression	0.13 (0.66)	0–6	1.01 (1.58)	0–8	3.14 (2.11)	0–10	1.14	7.44*
A priori cooperativeness	1.95 (0.22)	1–2	1.99 (0.12)	1–2	2.00 (0.00)	2–2	0.12	1.42
Consistency check	3.84 (1.45)	0–5	0.85 (1.23)	0–5	0.19 (0.43)	0–2	–0.75	–5.64*

Cohen's d and t scores were based on comparisons of low and high SIMS scorers in the patient sample from independent medical examinations

SRSI Self-Report Symptom Inventory, SIMS Structured Inventory of Malingered Symptomatology

* $p < .05$

Failure on one of the three parameters was treated as an indication of insufficient test effort. For the correlational analysis, scores on these three indicators were averaged.

Results Cronbach's alphas for the symptom and pseudosymptom scales were .94 and .91, respectively. On average, participants endorsed 26.45 SRSI genuine symptoms ($SD = 11.92$; range from 1 to 49) and 6.25 SRSI pseudosymptoms ($SD = 6.84$; range from 0 to 33). The mean SIMS score of the sample was 13.82 ($SD = 8.72$). A total of 69 participants (33.3 %) scored above the SIMS cutoff of 16, while 59 participants failed on the WMT (28.8 %). Means, standard deviations, and ranges for the SRSI scale scores are given in Table 2. All scale scores are separately presented for normal and high SIMS scorers in the sample. Both subgroups of independent medical examinations (IME) patients differed in all genuine symptom and pseudosymptom subscale scores as well as in the total scales scores. The highest effect size (Cohen's d) of all subscales was obtained for the cognitive pseudosymptom scale ($d = 1.84$) while effect sizes of the four other pseudosymptom scales varied from 0.98 to 1.20.

The correlation between SIMS and SRSI pseudosymptom scores was .83 in this sample, while the total SRSI genuine symptom scores also correlated significantly with both the SRSI pseudosymptoms ($r = .72$) and the SIMS ($r = .80$).

Furthermore, medium-sized significant correlations were observed between both the SRSI pseudosymptom and the SIMS scores and the WMT. The average score of IR, DR, and CNS correlated at $-.46$ with the SRSI pseudosymptom scale and at $-.45$ with the SIMS scores.

SRSI in Young Prison Inmates

Sample and Method Although the SRSI was primarily constructed for use in civil and social litigation contexts, its applicability in a criminal forensic context was tested in a master's thesis. Huhnt (2013) conducted a two-stage study with an initial sample of 65 inmates of a youth prison in Berlin, Germany. All participants were sentenced prisoners. They were all men and their mean age was 20.71 years ($SD = 1.73$). In a first test session, the sample was administered the SRSI. During a second test session 2 to 3 weeks later, the SIMS and a number of other instruments were administered to 45 participants.

Results The 65 participants endorsed on average 20.20 genuine symptoms ($SD = 9.50$) and 7.11 pseudosymptoms ($SD = 7.80$). Corresponding Cronbach's alphas for the two total scales were .90 and .92, respectively. Six of the 45 respondents (13.3 %) who were given the SIMS scored above the cut

score of >16 . The correlation between SIMS and SRSI pseudosymptoms was only .49 ($p < 0.05$), but this relatively moderate association might reflect the time passage between the first and the second session.

Reliability

Test-Retest Reliability Estimates

Samples and Method Retest reliability over a period of 14 days was checked with a combined sample of 30 healthy controls, half stemming from a Norwegian sample examined by Rafdal (2013) and half examined by Schlicht in the context of a German survey about laymen's popular beliefs and base rate estimates of malingering (Schlicht & Merten, 2014). The mean age of the 18 men and 12 women was 29.60 years ($SD = 10.86$; range: 21 to 61 years). The participants were asked to respond honestly to all questionnaires; the SIMS was also given to the 15 participants examined by Schlicht.

Results For the SRSI pseudosymptom scale, all participants scored at both sessions very low (with all scores < 6). The participants endorsed on average 9.33 genuine symptoms ($SD = 6.07$) and 0.90 pseudosymptoms ($SD = 1.56$). From first to second test session, there was a slight but significant drop by 1.43 genuine symptoms, $t(29) = 2.79$, $p < .05$, and 0.37 pseudosymptoms, $t(29) = 2.48$, $p < .05$).

Test-retest reliability (product-moment correlations) for the total SRSI genuine symptom scale was .91, with subscale reliabilities ranging from .69 for depressive symptoms to .94 for pain-related symptoms. Reliability for the total SRSI pseudosymptom scale was .87, with the subscale reliabilities ranging from .66 for anxiety/depression pseudosymptoms to 1.00 for motor pseudosymptoms. As no single item of the pain-related pseudosymptoms was endorsed by any of the participants at both test sessions, no reliability estimate was obtained for this subscale.

Internal Consistency Estimates

Samples and Method For an analysis of internal consistency at the subscale level, we combined all available German-language samples for the final SRSI version. The combined sample comprised $N = 387$ participants. For further details, see next section.

Results Table 3 contains both Cronbach's alpha estimates and split-half reliabilities. As can be seen, high alphas of .95 and .92 for the genuine symptom and the pseudosymptom total scales were obtained, with similar estimates for split-half

Table 3 Internal consistency scores (Cronbach's alphas; split half) for the pooled sample

Domain	Scale	Cronbach's α	Split-half reliability
Genuine symptom scales	Cognitive	.90	.89
	Depressive	.80	.76
	Pain	.89	.90
	Nonspecific somatic	.90	.92
	PTSD/anxiety	.81	.79
Pseudosymptom scales	Total symptoms	.95	.93
	Cognitive/memory	.80	.81
	Neurological: motor	.67	.61
	Neurological: sensory	.72	.70
	Pain	.75	.73
	Anxiety/depression	.79	.81
	Total pseudosymptoms	.92	.92
Consistency check		.85	–

$N = 387$

reliabilities. This indicates acceptable random error variance (Nunnally & Bernstein, 1994). Estimates for the separate 10-item subscales were lower, but all were in the acceptable range.

Final Analyses (Pooled Sample)

Samples and Method Data of the final SRSI version were available for a pooled sample of 387 German-speaking participants. These were the samples of the studies conducted by Giger (normative sample), Huhnt (young prison inmates), Stevens (independent medical examinations), and Schlicht (retest reliability data). The 265 men and 122 women had a mean age of 40.12 years ($SD = 13.75$). For 367 of them, SIMS protocols were available. As with the preliminary version, a receiver operating characteristic (ROC) analysis was performed on the pseudosymptom scores, using SIMS failure (> 16) as the criterion.

Results For the combined sample, SRSI genuine symptom and pseudosymptom scales correlated at .73. Both SRSI genuine symptom and pseudosymptom scores were highly associated with SIMS scores (with correlations of .82 for the two scales).

A total of 76 participants (20.71 % of $n = 367$) had a SIMS score exceeding the cutoff of 16. The ROC analysis of the total SRSI pseudosymptom scores yielded an area under the curve (AUC) of .931, with a standard error of measurement of .015 and a 95 % confidence interval reaching from .901 to .961. As was true for the results of

our preliminary analysis, this is an excellent accuracy. The ROC curve is presented in Fig. 1. Based on this ROC analysis, sensitivity and specificity estimates as well as likelihood ratios for all possible cutoffs are given in Table 4.

Discussion

We described psychometric data for the Self-Report Symptom Inventory, a new self-report measure that was designed to assess distorted symptom reporting. The SRSI was developed as an alternative to other free-standing tests, notably the SIMS (Smith, 2008; van Impelen et al., 2014). Unlike the SIMS, the SRSI includes both genuine and bogus symptoms. Moreover, it taps into psychopathological domains that are typically disputed in the civil arena (e.g., complaints of chronic pain, PTSD-like symptoms) and does not cover the feigning of extreme pathology that sometimes occurs in the criminal forensic arena (e.g., full-blown psychosis, mutism, disorientation, complete amnesia; e.g., Jaffe & Sharma, 1998).

The psychometric merits of the SRSI can be summarized as follows. First, a preliminary version of the SRSI was tested with a combined sample of $N=239$ and with these data, an empirical item selection was performed. The final SRSI pseudosymptom scale was found to correlate highly with the SIMS ($r=.81$), although the two instruments do not share the same items. A ROC analysis performed on the total pseudosymptom scores yielded an excellent AUC.

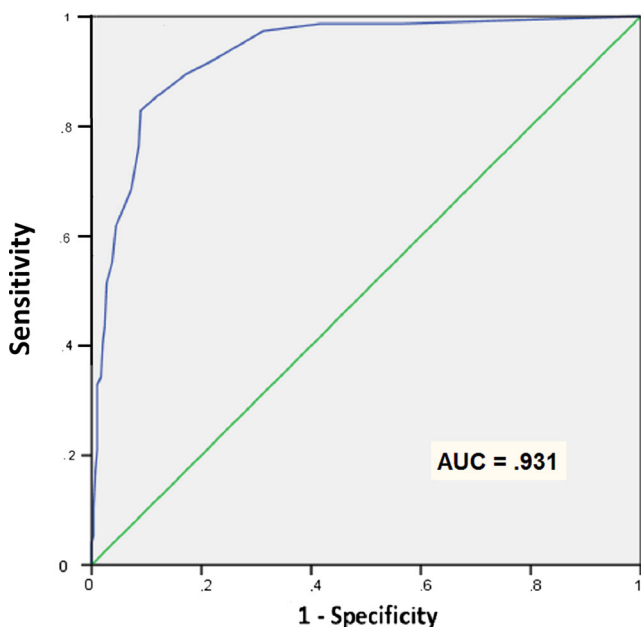


Fig. 1 Receiver operating characteristics of total SRSI pseudosymptom scores with SIMS failure as the criterion. Final analysis with a combined sample of $n=367$

Table 4 Results of receiver operating characteristic (ROC) analysis for SRSI pseudosymptom scale scores of the pooled sample

Cutoffs for SRSI pseudosymptom total	Sensitivity	Specificity	Likelihood ratio
>0	.987	.436	1.75
>1	.987	.584	2.37
>2	.974	.687	3.11
>3	.921	.777	4.13
>4	.895	.828	5.20
>5	.855	.880	7.13
>6	.829	.911	9.31
>7	.763	.914	8.87
>8	.684	.928	9.50
>9	.618	.955	13.73
>10	.553	.962	14.55
>11	.513	.973	19.00
>12	.434	.976	18.08
>13	.408	.979	19.43
>14	.342	.983	20.12
>15	.329	.990	32.90
>16	.289	.990	28.90
...			
>29	.039	1.00	–

$N=367$

SRSI Self-Report Symptom Inventory, SIMS Structured Inventory of Malingered Symptomatology

Second, for the pooled sample of general population participants, patients from independent medical examinations, and young prison inmates, pseudosymptom scores of the final SRSI version correlated highly with the SIMS ($r=.82$), thereby replicating the findings of the initial studies. Furthermore, SRSI pseudosymptoms correlated negatively with a PVT indexing underperformance (r with WMT = $-.45$), indicating that pseudosymptom endorsement on the SRSI tends to go hand in hand with lower effort on neuropsychological tests. However, correlations between distorted symptom report and PVTs are generally lower than those between conceptually similar self-report measures, which is understandable when one realizes that PVTs tap into a different aspect of response bias, namely underperformance. Thus, underperformance and distorted symptom report are loosely coupled dimensions (Merten & Merckelbach, 2013). Still, our results demonstrate that high scores on the SRSI pseudosymptom scale are associated with high SIMS scores and, to a lesser extent, underperformance. Overall, this pattern provides support to the validity of the SRSI pseudosymptom scale as an index of response distortion.

Third, data from the pooled sample yielded high Cronbach’s alphas and split-half reliabilities for the SRSI genuine symptom and pseudosymptom scales. Much the same

was true for their test-retest stability. On an individual scale level, alphas and split-half reliabilities can be considered as satisfactory bearing in mind the limited scale length of 10 items per subscale. Reliabilities tended to be lower for the pseudosymptom subscales.

Fourth, a ROC analysis performed on the pooled sample data for the final SRSI version resulted in an excellent area under the curve of .931 indicating a high degree of classification accuracy when the conventional cutoff of the German SIMS version is used as a criterion.

A number of limitations of the present results deserve comment. The data presented here are mainly focussed on comparisons of the SRSI pseudosymptom scores with SIMS scores. The SIMS has been identified as a suitable basis for the empirical item selection procedure. Both the SIMS and the pseudosymptom scales are composed of implausible, atypical, rare, or bogus symptoms. Consequently, the high correlations between the two instruments can be interpreted as estimates of concurrent validity. However, due to an equivalent construct validity of the two instruments, there is the danger of common-method variance. This may result in an over-estimation of the capability of the SRSI to identify distorted symptom report. As a consequence, the AUC obtained in the ROC analysis may also overestimate the SRSI's true capability to discriminate between distorted and non-distorted symptom endorsement. A related issue is that our analyses were based on a partial criterion (i.e., the SIMS) and did not include known-groups comparisons. Clearly, the research described here is just a first step and additional work needs to be done to evaluate the diagnostic merits of the SRSI with different methods and different instruments. Most importantly, genuine patient samples have to be studied; some of these studies are on its way. As one of the subscales is composed of pseudosymptoms related to anxiety and post-traumatic stress disorder (PTSD), one of the populations of interest will be genuine PTSD patients. Differentiating individuals with genuine PTSD from those who feign PTSD-related symptoms is an issue that has attracted much interest in the last 10 years or so (e.g., Howe, 2012; Young, 2014, for reviews). One concern is that some validity indicators have been shown to produce an elevated rate of misclassifying genuine PTSD patients (Lareau, 2011). Thus, future studies are required to examine the specificity of the SRSI pseudosymptom scales in patients with genuine PTSD.

Both the item selection procedures of the SRSI and many of the analyses presented in this paper were based on SIMS scores as an independent criterion. The SIMS was developed as a screening test for symptom exaggeration. Its limitations have been discussed in more detail by the authors elsewhere in detail (Merten et al., 2007; van Impelen et al., 2014). False-positive classifications are

likely to occur at a higher rate in patient populations with genuine severe psychopathology (e.g., Edens, Otto, & Dwyer, 1999). The same may be true for the SRSI, an issue that has to be carefully addressed in future studies. Arguably, the risk of false-positives depends upon the choice of an optimal cutoff, which was beyond the scope of the current article. With screening instruments, a recommended cutoff may allow for a pre-defined number of false-positives, usually 10 % or less. For instruments with intended higher diagnostic accuracy, the major focus will be minimizing false-positive classifications. The issue of choosing an optimal cutoff for the SRSI pseudosymptom scale will have to be addressed in future work.

Also, an in-depth analysis of the genuine symptom scales has not yet been performed. With future work, comparative data on these scales may become available from genuine patient samples enhancing the utility of the SRSI. The current analyses focused on the pseudosymptom scales, so the symptom scales can be seen, so far, as dummy items to disguise the test intention. For both the preliminary and the final test versions, high correlations between the pseudosymptom and the genuine symptom total scores were observed. These results need further investigation and clarification. It is well known that overreporting of potentially genuine symptoms and endorsement of atypical or bizarre symptoms (pseudosymptoms) constitute two related modes of responding in patients with feigned symptom presentation (e.g., Merten et al., 2007). Thus, a composite index appears promising on the basis of preliminary analyses, but this issue also needs to be taken up in future studies with large clinical samples.

Bearing these limitations in mind, we think that the SRSI may become a promising tool for detecting distorted symptom reports, alongside the SIMS. An alternative for the SIMS is needed whenever patients or evaluatees are familiar with the SIMS or might have been coached so that they can readily identify the SIMS as a SVT. In those circumstances, the SRSI might provide valuable information. The analyses presented in this paper are to be taken as a description of the first steps of a complex process of test construction and validation.

Acknowledgments Special thanks of the authors are due to Dr. Eva Friedel, who participated in the original conception of the instrument and the early phase of rational item construction. Furthermore, we acknowledge the contributions of Christina Fuchs (nee Diederich), Robert Lorenz, Sam Schlatow, Dr. Elisabeth Vossler-Thies, Lena Delattre, Maren Zon Rafald, Benno Huhnt, and David Schlicht for their assistance in the data collection phase.

Compliance with Ethical Standards

Conflict of Interest Prof. Stevens and Dr. Merten are regularly doing independent examinations in the fields of civil and social law. Prof. Merkelbach, Prof. Stevens, and Dr. Merten have participated in the development and adaptation of several validity tests.

References

- Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale, F (p). *Psychological Assessment*, 7, 424–431. doi:10.1037/1040-3590.7.4.424
- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21, 899–916. doi:10.1080/13825580600966391
- Bolinger, E., Reese, C., Suhr, J., & Larrabee, G. J. (2014). Susceptibility of the MMPI-2-RF Neurological Complaints and Cognitive Complaints scales to over-reporting in simulated head injury. *Archives of Clinical Neuropsychology*, 29, 7–15. doi:10.1093/arclin/act082
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence based approach*. New York: Guilford.
- Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, 7, 197–205. doi:10.1007/s12207-014-9198-7
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., . . . Silver, C. H. (National Academy of Neuropsychology Policy & Planning Committee). (2005). Symptom validity assessment: Practice issues and medical necessity. Official position of the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology*, 20, 419–426. doi:10.1016/j.acn.2005.02.002
- 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.
- Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability Consultative Exam: A new rating scale. *Archives of Clinical Neuropsychology*, 22, 1–14. doi:10.1016/j.acn.2006.10.003
- Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B., & Merckelbach, H. (2003). “Strukturierter Fragebogen Simulierter Symptome”. Die deutsche Version des “Structured Inventory of Malingered Symptomatology: SIMS” [The German version of the Structured Inventory of Malingered Symptomatology]. *Nervenarzt*, 74, 977–986. doi:10.1007/s00115-002-1438-5
- Dandachi-FitzGerald, B., Merten, T., Ponds, R. W. H. M., & Niemann, H. (2015). Europäische Umfrage zum Einsatz von Beschwerdewaldvalidierungstests: Ergebnisse der deutschen Teilnehmer [European survey on symptom validity assessment practices and beliefs: Results of the German participants]. *Zeitschrift für Neuropsychologie*, 26, 99–108. doi:10.1024/1016-264X/a000135
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28, 771–783. doi:10.1093/arclin/act073
- Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malingering psychopathology. *Journal of the American Academy of Psychiatry and Law*, 27, 387–396.
- Giger, P., & Merten, T. (2013). Swiss population-based reference data for six symptom validity tests. *Clinica y Salud*, 24, 153–159. doi:10.5093/cl2013a16
- Göbber, J., Petermann, F., Piegza, M., & Kobelt, A. (2012). Beschwerdewaldvalidierung bei Rehabilitanten mit Migrationshintergrund in der Psychosomatik [Symptom validity in patients with migration background in psychosomatic medicine]. *Die Rehabilitation*, 51, 356–364. doi:10.1055/s-0032-1323669
- Green, P. (2003). *Green’s word memory test. User’s manual*. Edmonton, Canada: Green’s Publishing.
- 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. doi:10.1080/13854040903155063
- Howe, L. L. S. (2012). Distinguishing genuine from malingered posttraumatic stress disorder in head injury litigation. In C. R. Reynolds & A. M. Horton (Eds.), *Detection of malingering during head injury litigation* (2nd ed., pp. 301–331). New York: Springer.
- Huhnt, B. (2013). *Stone cold manipulation or just an excuse? Psychopathy and blame attribution as possible explanations for faking symptoms in a forensic setting*. (Unpublished master’s thesis). Maastricht University, Faculty of Psychology and Neuroscience, Maastricht, The Netherlands.
- Jaffe, M. E., & Sharma, K. K. (1998). Malingering uncommon psychiatric symptoms among defendants charged under California’s Three Strikes and You’re Out law. *Journal of Forensic Sciences*, 43, 549–555. doi:10.1520/JFS16181J
- Lareau, C. R. (2011). Posttraumatic stress disorder and acute stress disorder. In D. Faust (Ed.), *Coping with psychiatric and psychological testimony* (6th ed., pp. 610–635). New York: Oxford University Press.
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 1–7. doi:10.1017/S1355617712000240
- Malcore, S. A., Schutte, C., Van Dyke, S. A., & Axelrod, B. N. (2015). The development of a reduced-item Structured Inventory of Malingered Symptomatology (SIMS). *Psychological Injury and Law*, 8, 95–99. doi:10.1007/s12207-015-9214-6
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists’ validity testing beliefs and practices: A survey on North American professionals. *The Clinical Neuropsychologist*, 29, 741–776. doi:10.1080/13854046.2015.1087597
- Meade, A. W., & Craig, B. C. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437–455. doi:10.1037/a0028085
- Merten, T., Friedel, E., & Stevens, A. (2007). Die Authentizität der Beschwerdewaldvalidierung in der neurologisch-psychiatrischen Begutachtung: eine Untersuchung mit dem Strukturierten Fragebogen Simulierter Symptome [Authenticity of symptom report in independent neurological and psychiatric examinations: A study with the Structured Inventory of Malingered Symptomatology]. *Praxis der Rechtspsychologie*, 17, 140–154.
- Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, 6, 122–137. doi:10.1007/s12207-013-9155-x
- Merten, T., Thies, E., Schneider, K., & Stevens, A. (2009). Symptom validity testing in claimants with alleged posttraumatic stress disorder: Comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingered Symptomatology, and the Word Memory Test. *Psychological Injury and Law*, 2, 284–293. doi:10.1007/s12207-009-9057-0
- Meyer, J. F., Faust, K. A., Faust, D., Baker, A. M., & Cook, N. E. (2013). Careless and random responding on clinical and research measures in the addictions: A concerning problem and investigation of their detection. *International Journal of Mental Health and Addiction*, 11, 292–306. doi:10.1007/s11469-012-9410-5

- Miller, H. A. (2001). *Miller-Forensic Assessment of Symptoms Test. Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morel, K. R. (2010). *Manual for the Morel Emotional Numbing Test for posttraumatic stress disorder* (2nd ed.). Author.
- Morey, L. C. (2007). *The Personality Assessment Inventory professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology*, *9*, 501–552. doi:10.1016/0887-6177(94)90041-8
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Rafdal, M. Z. (2013). *Norwegians never lie. A pilot study of the psychometric qualities of the Norwegian List of Indiscriminate Psychopathology*. (Unpublished master's thesis). Maastricht University, Faculty of Psychology and Neuroscience, Maastricht, The Netherlands.
- Rios, J., & Morey, L. C. (2013). Detecting feigned ADHD in later adolescence: An examination of three PAI-A negative distortion indicators. *Journal of Personality Assessment*, *95*, 594–599. doi:10.1080/00223891.2013.821071
- Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York: Guilford.
- Rogers, R., Gillard, N. D., Wooley, C. N., & Ross, C. A. (2012). The detection of feigned disabilities: The effectiveness of the Personality Assessment Inventory in a traumatized inpatient sample. *Assessment*, *19*, 77–88. doi:10.1177/1073191111422031
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*, 160–177. doi:10.1177/1073191103010002007
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms-2 (SIRS-2) and professional manual*. Lutz, FL: Psychological Assessment Resources.
- Santamaría Fernández, P. (2014). *Utilidad diagnóstica del inventario estructurado de simulación de síntomas (SIMS) en población española* [Diagnostic utility of the SIMS in the Spanish population.] (Unpublished doctoral thesis). Universidad Complutense de Madrid, Madrid, Spain.
- Schlicht, D., & Merten, T. (2014). Das Bild vorgetäuschter Gesundheitsstörungen in der öffentlichen Meinung [The picture of malingered symptom presentation in public opinion]. *Versicherungsmedizin*, *66*, 146–152.
- 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. doi:10.1136/jnnp.64.3.339
- Schroedter, J. H., Lechert, Y., & Lüttinger, P. (2006). *Die Umsetzung der Bildungsskala ISCED-1997 für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971 und die Mikrozensus 1976–2004. ZUMA-Methodenbericht 2006/08* [The adaptation of the educational scale ISCED-1997 for the census of 1970, the microcensus of 1971, and the microcensuses 1976 to 2004]. Mannheim, Germany: ZUMA.
- Schweizerisches Bundesamt für Statistik. (2011). *Statistisches Jahrbuch der Schweiz 2011* [Swiss statistical yearbook 2011]. Zurich, Switzerland: Neue Zürcher Zeitung.
- Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 323–342). New York: Guilford.
- United Nations Educational, Scientific and Cultural Organization. (1997). *International standard classification of education or ISCED-1997*. http://www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED_A.%20pdf. Download January 6, 2008.
- van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, *28*, 1336–1365. doi:10.1080/13854046.2014.984763
- Vitacco, M. J., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand trial patients: A known-groups comparison. *Law and Human Behavior*, *31*, 249–260. doi:10.1007/s10979-006-9062-8
- Vossler-Thies, E., Stevens, A., Engel, R. R., & Licha, C. (2013). Erfassung negative Antwortverzerrungen mit der deutschen Fassung des “Personality Assessment Inventory”, dem “Verhaltens- und Erlebensinventar” [Capturing negative response distortion with the German version of the Personality Assessment Inventory]. *Diagnostica*, *59*, 73–85. doi:10.1026/0012-1924/a000086
- Widows, M. R., & Smith, G. P. (2005). *SIMS—Structured Inventory of Malingered Symptomatology. Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Wiggins, C. W., Wygant, D. B., Hoelzle, J. B., & Gervais, R. O. (2012). The more you say the less it means: Overreporting and attenuated criterion validity in a forensic disability sample. *Psychological Injury and Law*, *5*, 162–173. doi:10.1007/s12207-012-9137-4
- Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court*. Dordrecht, NL: Springer.
- Zimmermann, P., Kowalski, J. T., Alliger-Horn, C., Danker-Hopfe, H., Engers, A., Meermann, R., & Hellweg, R. (2013). Detection of malingering in the assessment of occupational disability in the military. *German Journal of Psychiatry*, *16*, 54–60.