



Detecting Symptom Overreporting – Equivalence of the Dutch and German Self-Report Symptom Inventory

Brechje Dandachi-FitzGerald^{1,2} , Sinja Pienkohns¹, Thomas Merten³ , and Harald Merckelbach¹

¹Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands

²Faculty of Psychology, Open University, Heerlen, The Netherlands

³Department of Neurology, Vivantes Klinikum im Friedrichshain, Berlin, Germany

Abstract: The Self-Report Symptom Inventory (SRSI) intends to measure symptom overreporting. To assess the Dutch and German SRSI equivalence, both versions were split into two half-forms. Forty bilingual participants were randomly allocated to two groups that completed the first half in German and the second half in Dutch or vice versa. Each group completed the SRSI honestly and then under feigning instructions. For both conditions, the Dutch and German SRSI did not statistically significantly differ within and across the two groups. For most comparisons, the Bayes factor was ≥ 3 , indicating moderate evidence favoring the equivalence of language versions and half-forms. Genuine and pseudo-symptoms endorsement was significantly higher in the feigning than in the honest condition (both $Z_s = 5.44$, $r_{rb} = 1.00$). The SRSI standard cut score correctly identified honest responding and detected 80% of feigned responses. Our results align with Giger and Merten's (2019) German and French SRSI equivalence study.

Keywords: equivalence, symptom overreporting, feigning, Self-Report Symptom Inventory, Structured Inventory of Malingered Symptomatology



Patients may give incorrect answers to psychological tests for a variety of reasons, such as deliberate exaggeration in the pursuit of a particular benefit or responding carelessly to complete an assessment as quickly as possible (Merckelbach et al., 2019). Because invalid test data may compromise diagnostic conclusions and treatment recommendations with possible consequential harm (e.g., Van der Heide et al., 2020), it is essential to determine the validity of the psychological assessment before interpreting the data obtained. Based on a solid body of research, the current consensus is that the validity of test data should be determined with objective indicators, not just assumed or based on clinical judgment alone (Dandachi-FitzGerald & Martin, 2022). This approach is recommended in, for example, the recently updated consensus statement on validity assessment of the

American Academy of Clinical Neuropsychology (Sweet et al., 2021) and the guidelines on performance validity in neuropsychological assessments issued by the British Psychological Society (Moore et al., 2021).

To meet this recommendation in practice, sufficient evidence-based tools must be available to enable clinicians to gauge the validity of self-reported symptoms (with the so-called symptom validity tests; SVTs) and cognitive impairments (with the so-called performance validity tests; PVTs) with which individuals present. Meanwhile, there is an asymmetry in the extant literature with the number of available and well-researched instruments being far greater for PVTs than for SVTs (Giromini et al., 2022). When it comes to SVTs, the Structured Inventory of Malingered Symptomatology (SIMS; Shura et al., 2022; Smith & Burger, 1997; Widows & Smith, 2005; van Impelen et al., 2014) is the most commonly used stand-alone SVT (Dandachi-FitzGerald et al., 2013; Martin et al., 2015). The SIMS contains 75 items. Four of its five subscales examine the endorsement of rare, bizarre, or atypical symptomatology for mental disorders, such as

severe memory impairment or psychosis, while the fifth subscale – Depression – measures overreporting in the strict sense of the word. Although the psychometric track record of the SIMS is reasonable (for a recent overview, see Shura et al., 2022), it suffers from several limitations. For one, the SIMS was developed as a screening measure for evaluations in criminal procedures and therefore focuses on severe conditions, such as psychosis, amnesia, and mental retardation. The SIMS does not adequately cover conditions that are frequently encountered in compensations seeking and litigating contexts, such as concentration and mild or moderate memory difficulties, pain, or anxiety (Dandachi-FitzGerald et al., 2020; Hall & Poirier, 2021). Also, the SIMS mostly consists of a list of highly atypical or bizarre items, making it relatively easy to identify it as an SVT. Regardless of strengths and weaknesses of the SIMS, however, more SVTs are generally needed, for example, as alternative tests that can be included during repeated assessments and for tapping into symptom domains not covered by the SIMS scale structure.

The Self-Report Symptom Inventory

To expand the forensic expert's and the clinician's toolbox of SVTs and with the limitations of the SIMS in mind, Merten et al. (2016) developed a new, German-language test, dubbed the Self-Report Symptom Inventory (SRSI), which assesses noncredible reporting of less extreme and more common forms of psychopathology. The 107-item SRSI includes two warming-up items, five consistency-check items, 50 items addressing potentially genuine symptoms, and 50 items describing pseudosymptomatology. Thus, unlike the SIMS, the SRSI addresses both genuine symptoms and pseudosymptoms. Genuine symptom scales involve five subscales that include cognitive, depressive, pain, nonspecific somatic, and PTSD/anxiety symptoms. Similarly, pseudosymptom scales involve five subscales that gauge unlikely symptoms in the domains of cognitive functioning/memory, neurological motor complaints, neurological sensory complaints, pain, and PTSD/anxiety/depression symptoms. Each subscale consists of 10 items. Symptoms can be negated or affirmed, and affirmed answers are summed up so as to generate separate subscale scores and two total scores, one for genuine symptoms and one for pseudosymptoms.

Previous research generated four different cut scores to determine overreporting of pseudosymptoms as a function of the precise context in which the SRSI is employed (Merten et al., 2019): a liberal cut point (endorsing > 4 pseudosymptoms) for research purposes, a cut point for broad screening purposes (> 6), the cut point for standard diagnostic purposes (> 9), and a rigorous cut score (> 15).

Depending on the cut score, diagnostic determinations of overreporting can be made with different degrees of certainty. As a secondary validity indicator, the ratio between endorsed pseudosymptoms and endorsed genuine symptoms can be computed, reflecting differential endorsement of genuine complaints.

German validation studies found good psychometric properties for the SRSI (Merten et al., 2016, 2019, 2022). In the aggregated data set ($N = 367$) composed of healthy controls, experimental feigners, and patients, the sensitivity was .83 for the screening cut score (> 6) and .62 for the standard cut score (> 9), whereas the specificity was .91 (> 6) and .96 (> 9), respectively (Merten et al., 2019). In addition, the German version was found to possess high internal consistency (Cronbach's $\alpha = .95$ for the genuine symptoms and Cronbach's $\alpha = .92$ for pseudosymptoms) and adequate test-retest reliabilities ($r = .91$ for genuine symptoms and $r = .87$ for pseudosymptoms; Merten et al., 2016). Furthermore, pseudosymptoms scale score correlated substantially with the SIMS ($r = .82$). In several studies, PVTs were administered and correlations of SRSI pseudosymptoms with underperformance on these PVTs were in the small-to-medium size range (i.e., $r_s = -.13$ to $-.52$; Merten et al., 2019, 2022), indicating that overreporting tends to go hand in hand with underperformance. Note though that the modest correlations here do not suggest poor convergent validity because symptom and performance validity are two loosely coupled constructs (e.g., van Dyke et al., 2013).

Cross-Cultural Validation of the Dutch Version of the SRSI

The SRSI has been adapted into Dutch in subsequent steps of translation, back-translation, and fine-tuning of discrepancies. A number of studies showed its potential usefulness for Dutch clinical and forensic practice (e.g., Bošković et al., 2020; Merckelbach et al., 2018; van Helvoort et al., 2019). For instance, van Helvoort et al. (2019) examined the SRSI in a sample of 40 forensic patients admitted to a maximum security forensic psychiatric hospital. Participants were first instructed to fill in the SRSI honestly and subsequently in an exaggerated but convincing way. In the honest condition, two participants scored above the screening cut score (> 6) yielding a specificity of .95, and no participants scored above the standard cut score (> 9). Sensitivity ranged from .80 (> 9) to .92 (> 6). Merckelbach et al. (2018) collected data with the Dutch and English SRSI in 80 honest responders, 54 participants instructed to feign pain, and 53 participants instructed to feign anxiety. These authors found specificities of .92 (cut score > 6) to .97 (cut score > 9), whereas sensitivities ranged from .48 to .83 depending on the experimental

manipulation and the precise cut score that was used. These findings are promising, yet establishing cross-cultural equivalence is required for any foreign language version of an instrument. The meaning of an item might change in translation (e.g., proverbial expressions might be difficult or impossible to translate correctly), and many other cultural factors might unintentionally affect the interpretation and answering of items (for an overview, see Merten et al., in press). If a translated version lacks equivalence, it cannot be considered a valid replication of the original and should not be used in practice (International Test Commission, 2017). Thus, a study assessing the Dutch SRSI equivalence is essential for its use in Dutch practice. Moreover, establishing cross-cultural equivalence is highly relevant to research, allowing comparisons of studies examining the same concept (here, symptom overreporting) with different language versions of the same instrument in different populations (Franzen & European Consortium on Cross-Cultural Neuropsychology, 2021).

The Present Study

Using the paper of Giger and Merten (2019) on the equivalence of the French and German SRSI as an example, we tested the equivalence of the Dutch and German versions of the SRSI under conditions of honest reporting and feigning instructions in a sample of bilingual participants. These participants completed SRSI version of which one half consisted of Dutch and the other half of German items. Equivalence would be evident if and only if the two types of items generate similar patterns. Furthermore, we tested whether participants endorsed more genuine and more pseudosymptoms in the feigning than in the honest condition and whether endorsement of pseudosymptoms on the SRSI relates to SIMS scores in the expected direction (i.e., $r_s > .50$).

Method

Participants

We recruited adult participants from 18 to 65 years of age who were fluent in both Dutch and German languages, at least at level C1, implying good communication skills (Little, 2005). Participants were excluded if they (1) had a professional background in forensic assessments, (2) were intellectually disabled, and/or (3) suffered from severe psychopathology, such as acute psychotic symptoms, severe depression, or addiction. The inclusion and exclusion criteria were mentioned in the information letter and queried in the

demographical screening questionnaire (see the Measures section). With only these restrictions in place, a broad variety of genuine psychological symptoms is potentially possible. Recruitment took place via snowball sampling, starting with the first and second authors' bilingual contacts. The study was approved by the standing Ethics Review Committee of the Faculty of Psychology and Neuroscience of Maastricht University (ERCPN-230_132_11_2020).

In total, 42 persons participated in the study. Two participants had to be excluded from analyses, as one acknowledged to regularly use the German SRSI in practice and the second participant mentioned in the postexperimental manipulation check that he did not comply fully with the feigning instructions. Consequently, the final sample consisted of 40 participants (32 women, eight men) aged 21–61 years ($M = 41.4$, $SD = 12.9$). Three participants were bilingually raised and six participants had German as an acquired second language, while 31 participants had Dutch as an acquired second language. Of the whole sample, 30% ($n = 12$) reported currently living in Germany, and 70% in the Netherlands. The lowest educational level was the completion of 10th grade, and the majority had completed a master ($n = 16$) or bachelor ($n = 13$) program, and two participants held a PhD. Of all participants, seven were language teachers, only one was a psychologist, and the remaining participants had a large variety of professions, such as a musician, attorney, physicist, educator, ballet dance teacher, student, shop assistant, or office manager. In total, 27 participants (67.5%) indicated no prior mental or neurological disorders at all, while 13 participants indicated to have had or currently suffer from at least one condition. Participants mentioned to have suffered from head injury ($n = 3$), and depressive symptoms with varying intensity ($n = 5$), panic disorder ($n = 1$), post-traumatic stress disorder ($n = 1$), or unspecified psychopathological symptoms ($n = 3$). One participant reported to use psychopharmacological medication; no participant indicated substance abuse or dependency. Nine participants (22.5%) said they had been or currently were in psychotherapeutic treatment. Participants were randomly allocated to Group 1 ($n = 21$) or Group 2 ($n = 19$). The groups did not differ with regard to the frequency of psychiatric symptoms (zero vs. at least one) [$\chi^2(1) = 0.007$, $p = .94$] and experience with psychotherapy [$\chi^2(1) = 0.043$, $p = .84$].

Design

To facilitate comparisons across language version of the SRSI, the current study followed as closely as possible the equivalence study of Giger and Merten (2019) that contrasted the German and French SRSI in Swiss bilinguals.

We used the same inclusion criteria for bilingual individuals and relied on the same material (i.e., case vignette and role instruction, pre- and postexperimental questionnaires, demographical screening questionnaire, and item distribution of the SRSI half-forms; see below). The material, except from the SRSI item half-forms, is available at <https://osf.io/vn2zp/>.

The study relied on a 2 (between-subjects: Group 1 vs. Group 2) \times 2 (within-subjects: honest vs. feigning) design. The between-subject factor pertained to the order of German and Dutch SRSI items. Under both conditions, Group 1 received the first half of the SRSI in German and the second half in Dutch, while Group 2 had the reversed order. The within-subjects factor pertained to honest vs. feigning instructions (see below).

Measures

The Self-Report Symptom Inventory (SRSI)

The original German SRSI items (Merten et al., 2016) and their Dutch translations were used (Merckelbach et al., 2018). The SRSI contains 107 items, mostly addressing genuine symptoms (i.e., items of the following type: e.g., “I am feeling no interest in things”) or pseudosymptoms [i.e., items of the following type: e.g., “On a scale from 0 (*no headache*) to 10 (*maximum headache*), it is at 10 almost all the time”]. To ensure test security, the items presented here are for illustrative purposes only and do not represent actual items from the SRSI or SIMS. Five items assessing consistency were omitted because we wanted to create two halves with the same number of symptom items (Giger & Merten, 2019). Two items that assess cooperation (i.e., warming-up items) were always administered at the beginning of the questionnaire, leaving 100 items to be divided in two presumably equivalent half-forms. We used the same German half-forms as Giger and Merten (2019). In their study, items with similar psychometric properties were distributed in an even way to the two test halves. Each half-form consisted of an equal number of genuine symptoms and pseudo-symptoms, with each of the 10 subscales represented in an even way across the two halves.

The Structured Inventory of Malingered Symptomatology (SIMS)

The Dutch version of the SIMS was used as a concurrent measure of overreporting. The SIMS lists unlikely or bizarre symptoms (i.e., items of the following type: “There is a buzzing in my ears that keeps switching from left to right”) and has been translated into Dutch, with some item alterations to adapt its content to the Dutch culture (Merckelbach & Smith, 2003). For the Dutch version, a cut

score of > 16 was used, as previously recommended by Rogers et al. (1996). For this cut score, Dutch SIMS studies with instructed feigners ($N = 298$) found a sensitivity of .93 and a specificity of .98 (Merckelbach & Smith, 2003). In contrast, most American studies continue to use the original cut score of > 14 , as proposed in the SIMS manual (Widows & Smith, 2005).

Feigning Scenario

Participants completed the SRSI and the SIMS twice: the first time under the instructions to respond in an honest way to the questionnaires and the second time under the instruction to feign. Feigning instructions were introduced with a case vignette describing a person who had suffered a whiplash-like injury in an accident 6 months ago. To assess whether the injury had caused any residual impairments, the liability insurance company wanted the person to undergo a psychological assessment. Participants were instructed to assume the role of a person who felt entitled to financial compensation and, therefore, was motivated to convince the examiner that injury-related symptoms persisted to a disabling extent. The participants were not given any specific symptoms to feign but instead were free to choose the symptoms they wanted to feign.

Pre- and Postexperimental Checks

Two manipulation checks were administered to the participants before and after the feigning condition. The pre-experimental check was used to assess participants' understanding of their role instruction. If the questions were answered incorrectly, the scenario was once again presented. The postexperimental check items tested whether participants followed the scenario. Furthermore, these items queried the participants' strategies to fulfill the role requirements and asked for the symptoms that participants had attempted to feign.

Demographical Questionnaire

The demographic questionnaire consisted of items addressing the following topics: age, gender, level of education, current profession, and language proficiency. Additionally, a brief screening questionnaire was included to assess any prior or current neurological or mental conditions. The collected data were used to describe the sample and to ensure that the participants met the study's requirements.

Procedure

Due to the COVID-19 pandemic, testing took place online. Each participant was tested individually via

Zoom, a private, secure app for video conferences (Zoom Video Communications, Inc., 2021). The overall procedure took about one hour per session and was compensated for with 7.50 € or, in case of students, one course credit. To enhance motivation, participants were given the opportunity to enter a raffle to win an additional 50 € at the end of the study. At the beginning of the zoom call, participants were given a rough overview and were sent the link to the Qualtrics survey (Qualtrics, 2021), which they filled out while the first author was online for questions throughout their participation. Thus, all materials, including the SRSI and SIMS, were presented online as part of the Qualtrics survey. After obtaining informed consent, participants were randomly allocated to Group 1 (i.e., SRSI first half in German, second half in Dutch) or Group 2 (i.e., SRSI first half in Dutch, second half in German). First, a demographical screening questionnaire was presented (cf. supra). Then, participants were asked to fill out the SIMS and the half Dutch-half German SRSI in an honest way. Next, the participants read the case vignette and role instructions. Subsequently, they filled in the pre-experimental questionnaire, followed by the SIMS and the half Dutch-half German SRSI. Afterward, participants filled in the postexperimental questionnaire.

Data Analysis

Data were evaluated in several steps. First, to allow for comparisons with the equivalence study of Giger and Merten (2019), we applied null hypothesis significance testing to evaluate whether there were significant differences across and within groups (i.e., German vs. Dutch language and first vs. second test halves) for the genuine symptoms and pseudosymptoms total scores and for the honest and feigning conditions, separately. To this end, we used Mann-Whitney U tests because data were skewed. Given the priority of avoiding false negatives (Type II error) over false positives (type I error), an α level of .05 was used without correction for multiple testing.

Second, we used the Bayesian t test framework to quantify the evidence in favor of the equivalence of the Dutch and German SRSI (Linde et al., 2021). The equivalence Bayesian independent samples t test assumes normal distribution of the dependent variable. To determine whether we could use this test, we first performed both the Bayesian Mann-Whitney U test and the Bayesian standard t test for independent samples to quantify the evidence for the null hypothesis (i.e., there is no difference in language and half-forms for the genuine and pseudosymptom SRSI items) vs. the alternative hypothesis that

there is a significant difference (i.e., $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$; van Doorn et al., 2020). Because the outcomes of both tests were by and large similar for most comparisons, we proceeded with the equivalence Bayesian independent samples t test to quantify evidence for the interval-null hypothesis (i.e., the effect size falls within a specific interval) relative to evidence favoring the alternative hypothesis that the effect size falls outside the interval-null (i.e., $\delta \in I$ vs. $\delta \notin I$). The default interval range of -0.05 to 0.05 was used. For all analyses, δ was assigned a Cauchy prior distribution with $r = 1/\sqrt{2}$. The evidence is expressed in a Bayes factor. A BF of 1 indicates equal support for H_0 and H_1 (or for $\delta \in I$ and $\delta \notin I$ in case of equivalence testing). As a general rule of thumb, BFs between 1 and 3 are considered weak or anecdotal evidence, between 3 and 10 moderate evidence, and > 10 strong evidence (Van Doorn et al., 2021). Bayesian analysis was conducted with JASP (JASP Team, 2023). An annotated .jasp file is available at <https://osf.io/vn2zp/>.

Third, using Wilcoxon signed-rank tests with matched-pairs rank-biserial correlation (i.e., r_{rb}) as effect size (Kerby, 2014; van Doorn et al., 2020), we compared SRSI and SIMS scores of the honest and feigning conditions. We also looked at Spearman rank correlations between the two instruments. Fourth, classification accuracy parameters (i.e., sensitivity and specificity) were calculated for SRSI and SIMS.

Results

Equivalence of the Dutch and German SRSI

Neither in the honest condition (Table 1) nor in the feigning condition (Table 2) did the German and Dutch test halves generate significantly different scores. Similarly, comparisons of the half-forms within both language versions remained nonsignificant, and this was true for both genuine symptoms and pseudosymptoms in the two conditions. As can be seen in Table 3, for genuine symptoms in the honest condition, the evidence was in favor of equivalence of German and Dutch language versions, with BFs of 3.097 and 3.558 for the first and second half-forms, respectively. Thus, our data are roughly three times more likely to occur under the interval-null hypothesis than under the alternative (no equivalence) hypothesis. Evidence for the equivalence of the German and Dutch language pseudosymptom items in the honest condition was slightly weaker: BFs of 2.148 and 2.251 for the first and second half-forms, respectively, which means that the data are approximately twice more likely to occur under the interval-null hypothesis

Table 1. *M*, *SD*s, and Mann–Whitney *U* tests for the honest condition

SRSI	Genuine symptoms (<i>M</i> , <i>SD</i>)			Pseudosymptoms (<i>M</i> , <i>SD</i>)		
	1st half-form	2nd half-form	1st vs. 2nd half-form (<i>U</i> , <i>p</i>)	1st half-form	2nd half-form	1st vs. 2nd half-form (<i>U</i> , <i>p</i>)
German items	4.9 (3.5) ^a	4.1 (4.9) ^b	153.50 (.21)	0.4 (0.6) ^a	0.4 (1.1) ^b	171.00 (.31)
Dutch items	4.2 (3.6) ^b	3.9 (3.3) ^a	207.50 (.83)	0.7 (1.2) ^b	0.1 (0.4) ^a	240.00 (.13)
German vs. Dutch (<i>U</i> , <i>p</i>)	177.50 (.55)	182.00 (.63)		191.00 (.78)	193.50 (.79)	

Note. ^a*n* = 21. ^b*n* = 19.

Table 2. *M*, *SD*s, and Mann–Whitney *U* tests for the feigning condition

SRSI	Genuine symptoms (<i>M</i> , <i>SD</i>)			Pseudosymptoms (<i>M</i> , <i>SD</i>)		
	1st half-form	2nd half-form	1st vs. 2nd half-form (<i>U</i> , <i>p</i>)	1st half-form	2nd half-form	1st vs. 2nd half-form (<i>U</i> , <i>p</i>)
German items	19.10 (4.9) ^a	18.20 (4.6) ^b	166.50 (.37)	10.10 (6.8) ^a	11.10 (7.6) ^b	213.50 (.70)
Dutch items	18.90 (4.5) ^b	18.30 (5.2) ^a	209.50 (.79)	10.70 (7.2) ^b	12.20 (7.3) ^a	172.00 (.46)
German vs. Dutch (<i>U</i> , <i>p</i>)	185.50 (.70)	192.00 (.84)		197.00 (.95)	186.00 (.71)	

Note. ^a*n* = 21. ^b*n* = 19.

that under the alternative hypothesis. Regarding the comparison of the first half-form vs. second half-form of the Dutch SRSI pseudosymptoms, the Bayes factor was 0.720. This indicates that the data slightly favor the alternative (nonequivalence) hypothesis over the interval-null hypothesis, but the evidence for this is weak. In the feigning condition, the BF's favoring equivalence were ≥ 3 for all comparisons, which can be classified as moderate evidence for equivalence across language versions and half-forms (Table 4).

Given the equivalence of the Dutch and the German language versions, the scale scores of both half-forms were summed and subscale and total scale scores were computed. On average, participants in the honest condition endorsed 8.6 potentially genuine symptoms (*SD* = 7.3) and 0.8 pseudosymptoms (*SD* = 1.6). In the feigning condition, participants endorsed on average 37.3 (*SD* = 9.4) potentially genuine symptoms and 22.0 (*SD* = 13.9) pseudosymptoms.

SRSI Versus SIMS

Table 5 shows SRSI and SIMS scores in the honest and feigning conditions. Differences between the conditions were statistically significant (all *ps* < .001), indicating more endorsement of genuine symptoms and pseudosymptoms on the SRSI and higher SIMS total scores in the feigning condition compared with the honest condition. With the matched-pairs r_{tb} ranging between .956 and 1.00, the effect size was substantial (Kerby, 2014).

For the feigning condition, the correlations between the SIMS and the SRSI total genuine symptoms and pseudosymptoms were significant and, according to conventional standards, high: both $r_s = .88$ (95% CIs [0.78, 0.94]), *ps* < .01. Sensitivity and specificity were calculated for the various cut scores of the SIMS and the SRSI (see Table 6), demonstrating adequate specificity and sensitivity for all cut scores of the SIMS and the SRSI.

Symptoms Chosen to Feign

Participants' responses to the postexperimental questionnaire revealed the most popular symptoms in the feigning condition: concentration difficulties (97.5%), memory impairment (92.5%), psychological problems (72.5%), and pain (70.0%).

Discussion

We examined the equivalence of the German and Dutch versions of the SRSI. Our analyses yielded two main findings. First, we found no significant differences between language versions or half-forms of the SRSI. Second, Bayesian equivalence testing provided moderate evidence supporting the interval-null hypothesis for most comparisons, indicating that the language and half-form psychometric equivalence assumption was upheld. However, there were two exceptions to this general pattern.

Table 3. Equivalence Bayesian independent samples *t* test for the honest condition

Comparison	Model comparison	BF	Error %
German SRSI genuine symptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	2.846	8.658×10^{-6}
	$\delta \notin I$ vs. H_1	0.913	2.698×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.116	1.581×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.321	1.536×10^{-4}
Dutch SRSI genuine symptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	3.121	2.947×10^{-5}
	$\delta \notin I$ vs. H_1	0.900	1.022×10^{-4}
	$\delta \in I$ vs. $\delta \notin I$	3.468	5.306×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.288	6.380×10^{-4}
German vs. Dutch SRSI genuine symptoms for the 1st half-form	$\delta \in I$ vs. H_1	2.830	8.467×10^{-6}
	$\delta \notin I$ vs. H_1	0.914	2.622×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.097	1.548×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.323	1.484×10^{-4}
German vs. Dutch SRSI genuine symptoms for the 2nd half-form	$\delta \in I$ vs. H_1	3.191	1.081×10^{-5}
	$\delta \notin I$ vs. H_1	0.897	3.847×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.558	1.939×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.281	2.456×10^{-4}
German SRSI pseudosymptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	3.195	1.000×10^{-5}
	$\delta \notin I$ vs. H_1	0.897	3.564×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.564	1.794×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.281	2.278×10^{-4}
Dutch SRSI pseudosymptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	0.729	5.147×10^{-5}
	$\delta \notin I$ vs. H_1	1.013	3.703×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	0.720	1.043×10^{-4}
	$\delta \notin I$ vs. $\delta \in I$	1.390	5.397×10^{-5}
German vs. Dutch SRSI pseudosymptoms 1st half-form	$\delta \in I$ vs. H_1	2.132	5.283×10^{-6}
	$\delta \notin I$ vs. H_1	0.947	1.190×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	2.251	1.000×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.444	5.071×10^{-5}
German vs. Dutch SRSI pseudosymptoms 2nd half-form	$\delta \in I$ vs. H_1	2.043	5.413×10^{-6}
	$\delta \notin I$ vs. H_1	0.951	1.163×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	2.148	1.029×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.465	4.751×10^{-5}

Note. *I* ranges from -0.05 to 0.05 . BF = Bayes factor. Test for evidence favoring equivalence (i.e., $\delta \in I$ vs. $\delta \notin I$) in bold.

Specifically, we found no evidence for the equivalence of the Dutch first and second half-forms of pseudosymptom items in the honest condition, as opposed to moderate evidence for equivalence of these half-forms in the feigning condition. Additionally, the strength of evidence for equivalence between the Dutch and German SRSI pseudosymptoms was weaker (i.e., anecdotal) in the honest condition compared with the feigning condition (i.e., moderate). We believe that these exceptions can be attributed to the low endorsement of pseudosymptoms when participants responded honestly. Still, our findings suggest that the assumption of language and

half-form psychometric equivalence is supported for the overwhelming majority of the comparisons, with 15 of 16 comparisons showing evidence in favor of equivalence and most comparisons exhibiting moderate strength of evidence. Additional analyses with the bilingual SRSI revealed the expected pattern of significantly more endorsed genuine and pseudosymptoms in the feigning compared with the honest condition. Also, classification accuracy of the SRSI was high. At the standard cut score (i.e., > 9), the SRSI correctly classified 80% of the participants in the feigning conditions. None of the participants in the genuine condition was incorrectly classified as overreporting

Table 4. Equivalence Bayesian independent samples *t* test for the feigning condition

Comparison	Model comparison	BF	Error %
German SRSI genuine symptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	2.823	8.384×10^{-6}
	$\delta \notin I$ vs. H_1	0.914	2.589×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.088	1.533×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.324	1.462×10^{-4}
Dutch SRSI genuine symptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	3.031	2.185×10^{-6}
	$\delta \notin I$ vs. H_1	0.904	7.323×10^{-6}
	$\delta \in I$ vs. $\delta \notin I$	3.352	3.952×10^{-6}
	$\delta \notin I$ vs. $\delta \in I$	0.298	4.439×10^{-5}
German vs. Dutch SRSI genuine symptoms for the 1st half-form	$\delta \in I$ vs. H_1	3.199	1.052×10^{-5}
	$\delta \notin I$ vs. H_1	0.897	3.755×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.568	1.887×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.280	2.403×10^{-4}
German vs. Dutch SRSI genuine symptoms for the 2nd half-form	$\delta \in I$ vs. H_1	3.219	8.470×10^{-6}
	$\delta \notin I$ vs. H_1	0.896	3.044×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.594	1.517×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.278	1.959×10^{-4}
German SRSI pseudosymptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	3.004	3.962×10^{-6}
	$\delta \notin I$ vs. H_1	0.906	1.314×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.317	7.176×10^{-6}
	$\delta \notin I$ vs. $\delta \in I$	0.301	7.897×10^{-5}
Dutch SRSI pseudosymptoms 1st half-form vs. 2nd half-form	$\delta \in I$ vs. H_1	2.747	7.508×10^{-6}
	$\delta \notin I$ vs. H_1	0.918	2.247×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	2.993	1.378×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.334	1.234×10^{-4}
German vs. Dutch SRSI pseudosymptoms 1st half-form	$\delta \in I$ vs. H_1	3.133	2.591×10^{-5}
	$\delta \notin I$ vs. H_1	0.900	9.021×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.482	4.661×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.287	5.652×10^{-4}
German vs. Dutch SRSI pseudosymptoms 2nd half-form	$\delta \in I$ vs. H_1	2.960	1.009×10^{-5}
	$\delta \notin I$ vs. H_1	0.908	3.291×10^{-5}
	$\delta \in I$ vs. $\delta \notin I$	3.261	1.833×10^{-5}
	$\delta \notin I$ vs. $\delta \in I$	0.307	1.949×10^{-4}

Note. *I* ranges from -0.05 to 0.05 . BF = Bayes factor. Model testing evidence favoring equivalence (i.e., $\delta \in I$ vs. $\delta \notin I$) in bold.

symptoms. Overall, the constellation of findings supports the idea that the Dutch SRSI generates a similar pattern of data as the German original, thereby testifying to the equivalence of both test versions in detecting symptom overreporting.

This conclusion is further underlined when comparing the current results to those obtained in studies relying on the German and/or French SRSI. Thus, when comparing genuine and pseudosymptoms scores obtained with the German–French half-forms in the honest and feigning conditions (Giger & Merten, 2019; Tables 1 and 2) to our German–Dutch SRSI half-forms (Tables 1 and 2), similar

patterns of nonsignificance are evident. This is true for both comparisons of language versions and half-forms, with mostly a similar range of means, *SDs*, *U* statistics, and respective *p*-values. Similarly, as is true for the current study (Table 5), Giger and Merten (2019; Table 3) found in their feigning condition higher genuine and pseudosymptoms scores than in their honest condition. Importantly, the test statistics associated with these comparisons are highly comparable and, by all standards, substantial. Thus, the results of the two studies are strikingly similar. The fact that the findings of Giger and Merten (2019) and those of the present study are strongly converging has, to a certain extent,

Table 5. *M*, *SD*s, and Wilcoxon signed-rank tests for the honest vs. feigning condition (*N* = 40) for SRSI and SIMS

Symptom inventory	Honest condition		Feigning condition		<i>Z</i>	Rank-biserial correlation	95% CI for rank-biserial correlation	
	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range			LL	UL
SRSI genuine symptoms								
Cognitive	1.4 (1.8)	0–6	8.9 (2.4)	1–10	5.520***	.995	.990	.998
Depressive	1.1 (1.3)	0–5	5.8 (2.5)	0–9	5.314***	1.000	1.000	1.000
Pain	1.6 (1.9)	0–9	6.7 (3.6)	0–10	4.922***	.994	.987	.997
Nonspecific somatic	2.7 (2.6)	0–9	9.3 (1.2)	5–10	5.432***	.996	.992	.998
PTSD	1.8 (2.0)	0–7	6.5 (3.0)	1–10	5.217***	.956	.912	.979
Total	8.6 (7.3)	0–31	37.3 (9.4)	17–49	5.444***	1.000	1.000	1.000
Total Giger and Merten (2019) ^a	6.9 (6.4)	0–23	37.9 (8.1)	18–48	5.514***			
SRSI pseudosymptoms								
Cognitive	0.3 (0.6)	0–3	5.9 (2.8)	0–10	5.394***	1.000	1.000	1.000
Motor	0.1 (0.4)	0–2	3.6 (3.3)	0–10	4.871***	1.000	1.000	1.000
Sensory	0.3 (0.6)	0–3	3.6 (3.1)	0–10	5.012***	.982	.960	.991
Pain	0.1 (0.3)	0–1	4.4 (3.6)	0–10	4.794***	1.000	1.000	1.000
Mental	0.1 (0.4)	0–2	4.5 (3.5)	0–10	4.946***	1.000	1.000	1.000
Total	0.8 (1.6)	0–7	22.0 (13.9)	0–50	5.444***	1.000	1.000	1.000
Total Giger and Merten (2019) ^a	0.9 (1.1)	0–4	25.3 (13.5)	3–49	5.512***			
SRSI ratio								
Ratio	0.07 (0.12)	0.00–0.50	0.54 (0.26)	0.00–1.06	5.401***	.992	.984	.996
Ratio Giger and Merten (2019) ^a	0.17 (0.28)	0.00–1.00	0.63 (0.26)	0.12–1.05	4.960***			
SIMS								
SIMS	5.1 (3.1)	1–13	32.8 (13.6)	6–70	5.498***	.998	.995	.999
SIMS Giger and Merten (2019) ^{a,b}	3.8 (2.4)	0–9	32.1 (13.4)	0–54	5.512***			

Note. LL = lower limit. UL = upper limit. SIMS = Structured Inventory of Malingered Symptomatology. SRSI = Structured Inventory of Malingered Symptomatology. CI = confidence interval. ^aResults reported by Giger and Merten (2019); ^bPresent study used the Dutch SIMS, and Giger and Merten (2019) used the German SIMS. ****p* < .001.

Table 6. Sensitivity and specificity estimates for the SIMS and SRSI at different cut scores

Diagnostic accuracy statistic	SIMS		SRSI				
	(> 14)	(> 16)	Liberal (> 4)	Screening (> 6)	Standard (> 9)	Rigorous (> 15)	Ratio (> .288)
Sensitivity	.95	.90	.90	.85	.80	.60	.80
Specificity	1.0	1.0	.95	.95	1.00	1.00	.95

Note. SIMS = Structured Inventory of Malingered Symptomatology. SRSI = Self-Report Symptom Inventory.

to do with the design that both studies shared. Nevertheless, the samples were different, different language versions were employed, and the tests were administered online in the present study rather than in the traditional paper-and-pencil setup that Giger and Merten (2019) relied on. In sum, then, the results of both studies provide support for the cross-cultural equivalence of the SRSI language versions (see also Aryal et al., 2022, for similar results obtained with the English version of the SRSI).

An important caveat is that experimental analog studies such as the current one and that of Giger and Merten

(2019) overestimate classification accuracy of symptom and performance validity tests (e.g., Aryal et al., 2022; Vickery et al., 2001). On the one hand, sensitivity and specificity under real-world conditions, including forensic and clinical patient groups, are commonly lower than those obtained in well-planned and well-controlled experimental studies with their high degree of internal validity and their options to exclude noncompliant participants or control for potentially confounding factors. On the other hand, with their limitations in external validity in mind, analog studies are cost-effective for studying

selected aspects of interest, such as cross-cultural equivalence of tests. With data from field or archival studies, including those of forensic or clinical assessments, cross-language and cross-cultural comparisons are far more difficult to perform. Our findings provide support for the ability of the SRSI to discriminate between genuine and feigned symptom reporting, but the classification accuracy parameters that we found should be interpreted with caution. For any real-world use of the SRSI, the user should consult the estimates in the German SRSI manual (Merten et al., 2019). Thus, based on a multistudy analysis, this manual reported a sensitivity of .62 at the standard cut score, with a specificity of .96. Similarly, for the SIMS, more realistic classification statistics can be found in published meta-analyses and systematic reviews (Shura et al., 2022; van Impelen et al., 2014).

The case vignette that we used to instruct participants in the feigning condition alluded to injury-related symptoms but did not provide details about specific symptoms. As was true for the French–German SRSI equivalence study (Giger & Merten, 2019), feigning participants most frequently opted for concentration problems, memory problems, psychological symptoms, and pain. This list resonates with research showing that lay people have strong opinions about which symptoms lend themselves for feigning and which do not (e.g., Dandachi-FitzGerald et al., 2020; Peace & Masliuk, 2011). Interestingly, compared with the 48% of the Swiss sample of bilinguals in Giger and Merten's (2019) study, participants in our study more often (72.5%) opted to feign psychological symptoms, such as anxiety, depression, and trauma-related symptoms. Although this difference did not impact the study outcomes, it highlights that cultural differences might exist in which symptoms are considered to be beneficial for obtaining a certain benefit.

Limitations and Future Recommendations

The present study is only a first, albeit a promising, step in the validation of the Dutch SRSI version. Establishing psychometric equivalence requires further assessment of its validity and reliability, such as its construct validity or its test-retest reliability (Sousa & Rojjanasrirat, 2011). Arguably, the sample size in our study was quite small for an equivalence study and slightly uneven. Thus, to obtain robust evidence for equivalence, larger and more diverse samples consisting of symptomatic and nonsymptomatic individuals are needed. Also, for more fine-grained comparisons, preferably at the level of individual items, more participants are required. Nevertheless, the procedure of creating half-forms is considered adequate (Merten & Ruch, 1996) to establish language equivalence.

Considering that a sizeable minority in our sample reported suffering from genuine symptoms and/or undergoing treatment, we do think that our results can, to some extent, be generalized to the general population.

One obvious limitation is that we instructed participants in our feigning condition with only one scenario. To obtain more precise estimates of SRSI classificatory power under different conditions, various feigning instructions would be needed (Bošković et al., 2020; Merckelbach et al., 2009). Another limitation is that we presented the SRSI in the current study in a digital fashion. The SRSI was originally designed and validated as a paper-pencil instrument, raising caution about the online presentation mode employed in the present study (e.g., Merten et al., 2022). Nonetheless, the comparability of results to other studies that applied the SRSI as a paper-pencil instrument, specifically the equivalence study of Giger and Merten (2019), suggests a rather minimal influence of measurement format as long as specific measures to ensure compliance with the response conditions are taken.

As for future directions, more studies are needed with patients referred for clinical and forensic psychological assessment to further substantiate the construct validity and diagnostic acuity of the SRSI. In particular, known groups and differential prevalence study designs would further strengthen the evidence base for the SRSI. In experimental studies, case vignettes can be varied to study the sensitivity and specificity of the SRSI under different scenarios (e.g., Bošković et al., 2020). Also, spurred by the COVID-19 pandemic, studies are needed to explore the equivalence of online application modes (alone or with the clinician present in a secure video conference call) and the traditional in-person administration of the paper-pencil SRSI version (see Giromini et al., 2021, for an example with a different SVT). Future research on the cross-cultural validity of other available SRSI language versions (i.e., English, Spanish, Italian, Russian, Serbian, Norwegian, and Portuguese; Merten et al., 2022) is recommended. There is a need for more SVTs in clinical practice, and the empirical evidence so far supports the use of the SRSI as a suitable new instrument.

References

- Aryal, K., Merten, T., Akehurst, L., & Bošković, I. (2022). The English-language version of the Self-Report Symptom Inventory: A pilot analogue study with feigned head injury sequelae. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2022.2109158>
- Bošković, I., Merckelbach, H., Merten, T., Hope, L., & Jelčić, M. (2020). The Self-Report Symptom Inventory as an instrument for detecting symptom over-reporting: An exploratory study with

- instructed simulators. *European Journal of Psychological Assessment*, 36(5), 730–739. <https://doi.org/10.1027/1015-5759/a000547>
- Dandachi-FitzGerald, B., & Martin, P. K. (2022). Clinical judgement and clinically applied statistics: Description, benefits, and potential dangers when relying on either one individually in clinical practice. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 107–125). The Guilford Press.
- Dandachi-FitzGerald, B., Merckelbach, H., Bošković, I., & Jelacic, M. (2020). Do you know people who feign? Proxy respondents about feigned symptoms. *Psychological Injury and Law*, 13(3), 225–234. <https://doi.org/10.1007/s12207-020-09387-6>
- Dandachi-FitzGerald, B., Merckelbach, H., Merten, T., & Pienkows, S. (2023). *Detecting symptom over reporting: Equivalence of the Dutch and German* [Data set]. Self-Report Symptom Inventory. <https://osf.io/vn2zp/>
- 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(8), 771–783. <https://doi.org/10.1093/arclin/act073>
- Franzen, S. & European Consortium on Cross-Cultural Neuropsychology (2022). Cross-cultural neuropsychological assessment in Europe: Position statement of the European Consortium on Cross-Cultural Neuropsychology (ECCroN). *The Clinical Neuropsychologist*, 36(3), 546–557. <https://doi.org/10.1080/13854046.2021.1981456>
- Giger, P., & Merten, T. (2019). Equivalence of the German and the French versions of the Self-Report Symptom Inventory. *Swiss Journal of Psychology*, 78(1-2), 5–13. <https://doi.org/10.1024/1421-0185/A000218>
- Giromini, L., Pignolo, C., Young, G., Drogin, E. Y., Zennaro, A., & Vigliane, D. J. (2021). Comparability and validity of the online and in-person administrations of the Inventory of Problems-29. *Psychological Injury and Law*, 14(2), 77–88. <https://doi.org/10.1007/s12207-021-09406-0>
- Giromini, L., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: New articles, new issues. *Psychological Injury and Law*, 15(1), 1–21. <https://doi.org/10.1007/s12207-022-09444-2>
- Hall, H. V., & J., Poirier (Eds.) (2021). *Forensic psychology and neuropsychology for criminal and civil cases*. CRC Press.
- International Test Commission. (2017). ITC guidelines for translating and adapting tests (2nd ed.). *International Journal of Testing*, 18(2), 101–134. <https://doi.org/10.1080/15305058.2017.1398166>
- JASP Team. (2023). JASP (Version 0.17.1) [Computer software].
- Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, Article 1. <https://doi.org/10.2466/11.IT.3.1>
- Linde, M., Tendeiro, J. N., Selker, R., Wagenmakers, E. J., & van Ravenzwaaij, D. (2021). Decisions about equivalence: A comparison of TOST, HDI-ROPE, and the Bayes factor. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000402>
- Little, D. (2005). The common European framework and the European language portfolio: Involving learners and their judgments in the assessment process. *Language Testing*, 22(3), 321–336. <https://doi.org/10.1191/0265532205lt311oa>
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, 29(6), 741–776. <https://doi.org/10.1080/13854046.2015.1087597>
- Merckelbach, H., Dandachi-FitzGerald, B., Merten, T., & Bošković, I. (2018). De Self-Report Symptom Inventory (SRSI): Een instrument voor klachtenoverdrijving [The Self-Report Symptom Inventory (SRSI): An instrument for symptom exaggeration]. *De Psycholoog*, 53(3), 32–40. <https://doi.org/10.1007/s12207-016-9257-3>
- Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelacic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science*, 28(3), 321–326. <https://doi.org/10.1177/0963721419837681>
- Merckelbach, H., Smeets, T., & Jelacic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *The Journal of Forensic Psychiatry & Psychology*, 20(3), 378–386. <https://doi.org/10.1080/14789940802456686>
- Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, 18(2), 145–152. [https://doi.org/10.1016/S0887-6177\(01\)00191-3](https://doi.org/10.1016/S0887-6177(01)00191-3)
- Merten, T., Dandachi-FitzGerald, B., Bošković, I., Puente-López, E., & Merckelbach, H. (2022). The Self-Report Symptom Inventory. *Psychological Injury and Law*, 15(1), 94–103. <https://doi.org/10.1007/s12207-021-09434-w>
- Merten, T., Dandachi-FitzGerald, B., Puente-López, E., & Çetin, E. (in press). International perspectives on psychological injury and law: Cross-cultural aspects of symptom and performance validity assessment. In G. Young, T. Bailey, L. Giromini, L. Erdodi, R. Rogers, & B. Levitt (Eds.), *Handbook of psychological injury and law*. Springer Nature.
- Merten, T., Giger, P., Merckelbach, H., & Stevens, A. (2019). *Self-Report Symptom Inventory (SRSI): Handbuch* [SRSI professional manual]. Hogrefe.
- Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*, 9(2), 102–111. <https://doi.org/10.1007/s12207-016-9257-3>
- Merten, T., & Ruch, W. (1996). A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for depression. *Personality and Individual Differences*, 20(3), 281–291. [https://doi.org/10.1016/0191-8869\(95\)00185-9](https://doi.org/10.1016/0191-8869(95)00185-9)
- Moore, P., Bunnage, M., Kemp, S., Dorris, L., & Baker, G. (2021). *Guidance on the assessment of performance validity in neuropsychological assessment*. British Psychological Society.
- Peace, K. A., & Masliuk, K. A. (2011). Do motivations for malingering matter? Symptoms of malingered PTSD as a function of motivation and trauma type. *Psychological Injury and Law*, 4(1), 44–55. <https://doi.org/10.1007/s12207-011-9102-7>
- Qualtrics LLC. (2021). *Qualtrics* [Software]. <https://www.qualtrics.com/de/>
- Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment*, 67(2), 244–257. https://doi.org/10.1207/s15327752jpa6702_2
- Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured Inventory of Malingered Symptomatology: A psychometric review. *Psychological Injury and Law*, 15(1), 64–78. <https://doi.org/10.1007/s12207-021-09432-y>
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *The Journal of the American Academy of Psychiatry and the Law*, 25(2), 183–189. <https://doi.org/10.1007/s12207-021-09432-y>
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline.

- Journal of Evaluation in Clinical Practice*, 17(2), 268–274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 35(6), 1053–1106. <https://doi.org/10.1080/13854046.2021.1896036>
- van der Heide, D., Bošković, I., van Harten, P., & Merckelbach, H. (2020). Overlooking feigning behavior may result in potential harmful treatment interventions: Two cases of undetected malingering. *Journal of Forensic Sciences*, 65(4), 1371–1375. <https://doi.org/10.1111/1556-4029.14320>
- van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's ρ . *Journal of Applied Statistics*, 47(16), 2984–3006. <https://doi.org/10.1080/02664763.2019.1709053>
- van Doorn, J., van den BerghBöhm, D. U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. M., Sarafoglou, A., Stefan, A., Völkel, A. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin and Review*, 28(3), 813–826. <https://doi.org/10.3758/s13423-020-01798-5>
- Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist*, 27(8), 1234–1246. <https://doi.org/10.1080/13854046.2013.835447>
- van Helvoort, D., Merckelbach, H., & Merten, T. (2019). The Self-Report Symptom Inventory (SRSI) is sensitive to instructed feigning, but not to genuine psychopathology in male forensic inpatients: An initial study. *The Clinical Neuropsychologist*, 33(6), 1069–1082. <https://doi.org/10.1080/13854046.2018.1559359>
- 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(8), 1336–1365. <https://doi.org/10.1080/13854046.2014.984763>
- Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, 16(1), 45–73. [https://doi.org/10.1016/S0887-6177\(99\)00058-X](https://doi.org/10.1016/S0887-6177(99)00058-X)
- Widows, M. R., & Smith, G. P. (2005). *SIMS – Structured Inventory of malingered symptomatology. Professional manual*. Psychological Assessment Resources.
- Zoom Video Communications, Inc. (2021). *Zoom-Cloudmeetings* (Version: 5.7.1) [Computer app]. https://zoom.us/download#room_client

History

Received October 26, 2022

Revision received March 6, 2023

Accepted May 3, 2023

Published online June 20, 2023

Section: Clinical Psychology

Conflict of Interest

There was no third-party funding for the article. Interest to disclose: T. M. and H. M. are among the authors of the German-language Self-Report Symptom Inventory, and B. DF., T. M., and H. M. are the authors of its Dutch-language version. The test is commercially distributed by Hogrefe Publishers.

Publication Ethics

Informed consent was obtained from all participants included in the study. The study was approved by the standing Ethics Review Committee of the Faculty of Psychology and Neuroscience of Maastricht University (ERC PN-230_132_11_2020).

Open Science

Open Data: The information needed to reproduce all of the reported results is available at <https://osf.io/vn2zp/> (Dandachi-FitzGerald et al., 2023).

Open Materials: The information needed to reproduce all of the reported methodology is available at <https://osf.io/vn2zp/> (Dandachi-FitzGerald et al., 2023).

ORCID

Brechje Dandachi-FitzGerald

 <https://orcid.org/0000-0002-8984-8192>

Thomas Merten

 <https://orcid.org/0000-0001-8423-8507>

Brechje Dandachi-FitzGerald

Faculty of Psychology and Neuroscience

Maastricht University

PO Box 616

6200 MD Maastricht

The Netherlands

b.fitzgerald@maastrichtuniversity.nl