Psychological Injury and Law
https://doi.org/10.1007/s12207-025-09543-w

ORIGINAL RESEARCH

Sensitivity of the Genuine Symptoms Scale of the Self-Report Symptom Inventory (SRSI) to Psychopathology: Enhancing the Informational Value of a Symptom Validity Test for Symptom Overreporting

Louis De Page[1] · **Brechje Dandachi-FitzGerald**[2,3] · **Harald Merckelbach**[3]

Received: 28 February 2025 / Accepted: 4 July 2025

Abstract
The Self-Report Symptom Inventory (SRSI) is primarily utilized to identify symptom overreporting. To make this purpose less apparent, the SRSI includes a blend of 50 genuine symptoms and 50 unlikely symptoms (i.e., pseudosymptoms) presented in a mixed format. Studies have shown that the pseudosymptoms scale is effective in detecting symptom overreporting, but the genuine symptom items of the SRSI have typically been regarded as filler content. Our study aimed to determine whether these genuine symptom items could yield clinically meaningful information once overreporting, underreporting, and inattentive/random responding have been screened out, recognizing the limitations imposed by a restricted set of screening indices. We analyzed SRSI genuine symptoms scores in a screened sample of 100 psychotherapy patients and 81 job applicants. Psychotherapy patients endorsed significantly more genuine symptoms than job applicants, and a receiver operating characteristics analysis of the genuine symptoms main scale yielded an AUC of 0.94 (95% CI [.89, .97]), indicating strong discrimination. Subscale AUCs were generally below .90, suggesting that the SRSI genuine symptoms main scale, rather than its subscales, may serve as a useful tool for identifying cases that may require clinical attention. Additionally, we found a strong correlation ($r = .78$, $p < .001$) between the SRSI genuine symptoms main scale and the higher-order Emotional/Internalizing Dysfunction (EID) scale of the MMPI-2-RF, further highlighting the clinical value of the genuine symptoms main scale.

Keywords Self-report symptom inventory · SRSI · Symptom validity · Overreporting · Genuine symptoms · MMPI · Supernormality

Introduction

In recent decades, researchers have crafted tools to help clinicians spot when patients might be exaggerating their symptoms during diagnostic evaluations (see for reviews, e.g., Giromini et al., 2022a, b; Kirk et al., 2020; Lippa, 2018). These tools are often referred to as validity tests, encompassing both symptom validity tests and performance validity tests (Larrabee, 2012). Although such tests are mostly rather effective in screening for symptom exaggeration, clinicians might be reluctant to use them. For example, Dandachi-FitzGerald et al. (2013) noticed in their survey among European neuropsychologists ($N = 515$) that a sizeable minority indicated to never use an instrument to check for symptom exaggeration. More specifically, 17% said not to employ such instruments in clinical assessments, whereas almost 10% said to refrain from using these instruments in forensic settings. Other researchers observed an even more widespread tendency to dispense with symptom validity tests. Thus, in their survey among neuropsychologists ($N = 588$) in the UK, McCarter et al. (2009; p. 1054) concluded that "66% of clinical workers and 13% of medico-legal workers [reported that they] never use any formal

✉ Louis De Page
Louis.de.page@gmail.com

[1] Faculty of Psychological & Educational Sciences, Vrije Universiteit Brussel, Brussels, Belgium

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

[3] Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

Published online: 22 September 2025

symptom validity test or procedure to validate their impression." Surveys that are more recent reveal a similar trend (e.g., Giromini et al., 2022a, b; but see Martin et al., 2024). For instance, Uiterwijk et al. (2021) conducted a survey of 102 Australian psychologists performing neuropsychological assessments and found that these professionals utilized symptom validity tests much less frequently in clinical environments compared to medico-legal evaluations.

Why would clinical professionals hesitate to use symptom validity tests when many of these instruments possess acceptable sensitivity and specificity and might assist the professional in screening for symptom exaggeration? It would be too easy to attribute this hesitance to a naïve lack of appreciation or even wrong ideas about the psychometrics of symptom validity tests. McCarter et al. (2009) did ask their sample of professionals what justified their choice *not* to administer symptom validity tests to their patients. Survey participants frequently expressed concerns that appeared understandable. For example, a nontrivial minority ($n = 35$; 27%) argued that symptom validity tests require additional time that is not always available during diagnostic assessments (for a similar observation, see Uiterwijk et al., 2021). Others ($n = 9$; 7%) indicated that symptom validity tests may compromise the relationship between clinicians and patients.

The first point gains in strength when one considers that the base rate of symptom overreporting is relatively low in clinical settings, with some experts estimating that it gravitates in the $15\% \pm 15\%$ range (Young, 2015). The second point is a valid consideration given that symptom validity tests such as the widely used Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997) often involve bogus items that strain credulity to an extent that patients might feel uneasy (e.g., "the capital of Italy is Hungary"). Here, symptom validity test items could induce reactance in patients, thereby modifying the very phenomenon that clinicians want to measure (i.e., symptom levels; French & Sutton, 2010). Importantly, researchers in symptom validity testing have yet to explore whether bizarre items might elicit reactive attitudes in patients. Given this context, it is understandable that some clinicians trust their observations, instincts, and overall clinical judgment more than formal symptom validity testing.

Some symptom validity tests could potentially address these concerns, one example being the Self-Report Symptom Inventory (SRSI; Merten et al., 2016, 2022). This 107-item tool includes a blend of pseudosymptoms and genuine symptoms, addressing issues that are relatively common (e.g., anxiety, depression, functional neurological symptoms). Both types of symptoms are represented with 50 items, and it is the alternation between pseudosymptoms and genuine symptoms that makes the SRSI less obviously recognizable as a symptom validity test. To check whether someone engages in overreporting symptoms, the clinician focuses on the number of pseudosymptoms that this person endorses: if that number exceeds a predetermined cutoff (e.g., 6 for screening purposes and 9 for diagnostic assessments), the individual apparently exaggerates symptom levels. Numerous clinical and experimental studies support this conclusion with a fair degree of confidence (e.g., Aryal et al., 2022; Dandachi-FitzGerald et al., 2023a; Merten et al., 2020). False positive rates typically remain around 5%, while receiver operating characteristics (ROC) analyses of the SRSI pseudosymptoms scale generates areas under the curve (AUCs) between .90 and .96 (see also Merten et al., 2022).

Research on the psychometric properties of the SRSI has primarily focused on the pseudosymptoms scale and its effectiveness in detecting symptom overreporting. Notably, substantial efforts have been made to evaluate the equivalence of the French, Dutch, and Italian translations of the SRSI, yielding largely promising results (e.g., Dandachi-FitzGerald et al., 2023b; Geurten et al., 2018; Giger & Merten, 2019; Ribatti et al., 2024). However, in this type of study, the items on the genuine symptoms scale are simply used as filler, resulting in limited understanding of whether the genuine symptom scale of the SRSI could provide valuable diagnostic information. For instance, if someone scores below the cutoff for the pseudosymptoms (i.e., < 6 or < 9), does their endorsement of items from the genuine symptoms scale provide a reliable indication of whether they are experiencing clinically significant levels of distress? If so, this could make the SRSI more appealing to clinicians, given the concerns we previously discussed.

Aim of the Study

This study aimed to explore the utility of the SRSI beyond identifying symptom overreporting, focusing specifically on the genuine symptoms scale's potential to flag serious mental health problems. Briefly, we administered the SRSI along with other tests to two groups and contrasted their genuine symptom scores: outpatients who were in psychotherapeutic treatment for their mental health problems, and job applicants who underwent a diagnostic routine assessment. We opted for job applicants because they are, a priori, expected to suffer less frequently from serious mental health problems and have no motivation to exaggerate symptoms, making them a suitable comparison group for evaluating the sensitivity of the SRSI genuine symptoms scale. Additionally, we wanted to keep test administration as comparable as possible across groups, so we included patients and job applicants from one institution (see below), where the SRSI is part of the standard routine. We tested whether endorsement of genuine SRSI symptoms could differentiate between these two groups to such degree that it would allow clinicians to employ the SRSI as an instrument to screen for mental

health problems in people with unremarkable scores on the pseudosymptoms scale.

Method

Participants

The study was part of an ongoing research project at Mediter Center, a facility specialized in (neuro)psychological assessment and outpatient psychotherapy, located in Halle, Belgium. The project was approved by the standing ethical committee of the Vrij Universiteit Brussel (Ref 2020–527). Participants ($N = 436$) underwent psychodiagnostic assessment at Mediter Center, either because they were referred for psychotherapy ($n = 145$; 72 women) or they were referred in the context of a job application procedure ($n = 291$; 35 women). During the assessments, participants completed several psychological tests and tasks. However, the present study focuses specifically on the SRSI, the Supernormality Scale (SS; Cima et al., 2003), and the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011). The test batteries varied between groups. Job applicants underwent a standardized assessment battery as part of a routine psychological screening procedure, conducted under direct supervision of a senior psychologist in a controlled testing environment and lasting approximately four hours. In contrast, psychotherapy patients completed their assessments during the initial therapy sessions, with test selection tailored to the referral questions and the individual's presenting problems. Most patients sought help for internalizing disorders such as depression, anxiety, and other emotional dysregulation issues (e.g., burnout and relational difficulties). A subgroup, however, had been referred for more in-depth assessment and treatment following unsuccessful previous interventions. For these individuals, the assessment focused on exploring defensive or resistant patterns of functioning that might hinder therapeutic progress.

The sample included both Dutch ($n = 204$) and French speaking ($n = 232$) participants, who were administered the Dutch or French test versions, respectively. In a previous study, we found that the Dutch and French versions of the SRSI are psychometrically equivalent (Dandachi-FitzGerald et al., 2023a; see also Geurten et al., 2018). Similarly, there are good reasons to believe that the French version of the SS is highly comparable to its Dutch counterpart (De Page & Merckelbach, 2021; De Page et al., 2013). The Dutch/Flemish and French versions of the MMPI-2 have been well-studied and based on numerous studies, we may assume that they produce similar normative data (e.g., Butcher et al., 2003). The same holds by implication for the MMPI-2-RF, as this version was directly derived from the more extensive MMPI-2 (see also Sleep et al., 2015).

With this in mind, we collapsed the data of Dutch and French speaking participants. The current sample partially overlaps with the one described in the Dandachi-FitzGerald et al. (2023a) paper, but the overlap is not complete. Compared to the Dandachi-FitzGerald et al. (2023a) sample, 122 of the 436 records (28%) were new and not represented in the previous study. Most importantly, while the previous study addressed participants' pseudosymptoms, the current study targets the genuine symptoms scale of the SRSI.

As expected, individuals in the psychotherapy group were older than those in the job applicant group, with mean ages of 39.18 years ($SD = 12.08$) and 23.09 years ($SD = 7.65$), respectively; $t(433) = 16.90$, $p < .01$. The psychotherapy group was more gender-balanced than the job applicants: 49% vs. 12% women, respectively, Fisher's exact $p < .001$. There were less French speakers in the psychotherapy group than in the job applicant group (i.e.., 46% vs 55%), Fisher's exact $p < .05$.

Instruments

Self-Report Symptom Inventory (SRSI)

The SRSI (Merten et al., 2016, 2019, 2022) consists of 107 items with a true-or-false response format. Seven items test for cooperation and response consistency. The 100 remaining items allude to symptoms and health problems. Items are easy to understand. During the item construction and selection process of the original SRSI, experienced clinicians and researchers with expertise in the domain of validity assessment placed much emphasis on readability and clarity of item content (Merten et al., 2019). Thus, the SRSI does not place high demands on reading ability. Recent research in the Netherlands tested the SRSI among children around the age of 12, and readability did not pose any issues for them (Dandachi-FitzGerald et al., 2024).

The 100 symptom items of the SRSI fall into either of two main scales of 50 items each. One main scale, which is the focus of the current study, consists of potentially plausible complaints (i.e., genuine symptoms, *Xx*; e.g., "Feeling no interest in things"). The other main scale comprises less credible symptoms (i.e., pseudosymptoms, *Yy*; e.g., "On a scale from 0 (no headache) to 10 (maximum headache), my headaches are at '10' almost all the time").

The two main scales each include five subscales, with every subscale consisting of 10 items that cover either genuine or pseudosymptoms in specific domains. Below, we focus on the genuine symptoms main scale and its subscales, which are as follows: cognitive problems (*Xco*), depression (*XDe*), pain (*XPa*), nonspecific somatic complaints (*XSo*), and anxiety (including PTSD; *XAn*). The range of subscale

scores (i.e., sum of endorsed symptoms) varies between 0 and 10. Accordingly, total scores for each of the two main scales range from 0 to 50. Previous studies have reported total scores on the genuine symptoms scale (i.e., *Xx*) across various samples, with mean scores of 9.8 ($SD = 8.0$) in a non-screened heterogeneous population sample (Ribatti et al., 2025), 16.4 ($SD = 8.6$) in memory clinic outpatients with mild cognitive impairment (Czornik et al., 2021), and 25.2 ($SD = 10.9$) in psychosomatic rehabilitation clinic inpatients (Merten et al., 2020).

For the pseudosymptoms scale (i.e., *Yy*), Merten et al. (2016) recommended > 9 as the most optimal cut point in standard diagnostic settings (with a maximum of 5% false positives). In the current study, Cronbach alpha's for the genuine and pseudosymptoms main scales were .89 and .88, respectively.

Supernormality Scale (SS)

The SS is a 37 true–false questionnaire developed by Cima et al. (2003) to screen for the tendency to deny common symptoms, i.e., symptoms that most people have from time to time (e.g., feeling sad, having intrusions about emotional events). Apart from five distractor items, the SS comprises two scales: social desirability (11 items; e.g., "I try to help everybody who has problems") and supernormality (21 items; e.g., "My mental state is completely normal"). After correction for reversed coding and excluding distractor items, yes-answers are summed to obtain total SS scores (range 0–32). Cima et al. (2003) concluded that the psychometric features of the SS are acceptable and proposed several cut points. In the current study, we employed a cutoff of > 22 to obtain the best trade-off between excluding participants who underreport symptoms (sensitivity) and including participants who do not underreport symptoms (specificity). Cronbach's alpha for the total SS scale was .81.

Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF)

The MMPI-2 (Butcher et al., 1989) is a widely used psychological assessment tool designed to evaluate personality structure and psychopathology. Comprising 567 true–false questions, it assesses various psychological conditions and personality dimensions across ten clinical scales and various subscales. The MMPI-2 is employed in clinical, forensic, and research contexts to diagnose mental disorders, inform treatment planning, and evaluate the psychological fitness of individuals for particular roles. In 2008, Ben-Porath and Tellegen (2008/2011) developed the MMPI-2-RF form using items from the MMPI-2, introducing new clinical, validity, and specific problem scales that exhibit improved psychometric properties, particularly without any item overlap. We derived MMPI-2-RF scores from participants' MMPI-2 responses, following the procedure established by Van der Heijden et al. (2010). In the current study, we were primarily interested in the following MMPI-2(-RF) validity parameters: CNS (cannot say-responses), VRIN-r (variable response inconsistency), and TRIN-r (true response inconsistency) as indices of inattentive or random responding, and L-r (uncommon virtues) and F-r (infrequent responses) as indices of underreporting and overreporting, respectively (Sleep et al., 2015; Ingram & Ternes, 2016; Sharf et al., 2017). In screening participants, we employed the following MMPI-2(-RF) cutoffs: CNS > 30,[1] VRIN-r, TRIN-r, and L-r ≥ 80, and F-r ≥ 100 (Ben-Porath & Tellegen, 2008/2011).

We opted for F-r from the family of F scales because this index is thought to be particularly sensitive to exaggeration of mental health symptoms rather than medical symptoms. Also, it has been previously used in research on the SRSI (e.g., Merten et al., 2022), facilitating comparisons across studies. Aparcero et al. (2022; p. 757) concluded in their meta-analysis that: "When analyzed in terms of actual scale scores, these data indicate that non-genuine responders reporting exaggerated or feigned psychiatric symptoms typically generate F/Fr and Fp/Fp-r T-scores exceeding 100. Although many genuine psychiatric patients also obtain elevated scores on these scales, T-scores above 95 on the F and F-r scales and above 75 on the Fp and Fp-r scales (i.e., two standard deviations above the mean, or $M + 2\ SD$; Sharf et al., 2017) are rare and are likely to indicate feigning." Still, a cutoff of ≥ 100T might be relatively liberal (Burchett & Bagby, 2022; Sharf et al., 2017) in the sense that it may not meet the 90% specificity criterion. However, our primary goal was to maximize sensitivity—i.e., ruling out individuals who overreport symptoms. We also acknowledge that, considering accumulating meta-analytic findings, it would have been conceivable to include additional indices from the F family (e.g., FBS-r, Fp-r). However, due to the potential collinearity this would introduce, we chose a parsimonious approach, focusing on F-r as a screener for overreporting of mental health problems, alongside the pseudosymptoms scale of the SRSI (but see below).

Additionally, we were interested in the higher-order EID (Emotional/Internalizing Dysfunction) scale and the five somatic/cognitive specific problem scales of the MMPI-2-RF because the genuine symptoms subscale of the SRSI aims to cover these domains. EID is a broad measure of psychopathology, though it is particularly sensitive to internalizing disorders, making it similar to the RCd measure of general distress (Greene, 2011). EID and the five somatic/

[1] We applied the MMPI-2 criterion because participants completed the full MMPI-2, and the CNS scale is specifically designed to account for the test's entire length.

cognitive specific problem scales of the MMPI-2-RF do not overlap. Low EID scores are predictive of the absence of a psychiatric diagnosis (e.g., Haber & Baum, 2014). Following the suggestion of Romero et al. (2017), we anticipated that SRSI genuine symptoms would align most closely with the internalizing domain (see also Menton, 2022; Sellbom et al., 2012).

Procedure

As part of the clinical routine, psychotherapy patients were assessed during their first session, prior to the commencement of treatment. The instruments employed during psychological testing were dependent on the referral question, but as a rule the SRSI, SS, and MMPI-2(-RF) were included to screen for symptom overreporting, symptom underreporting, and inattentive or random responding. This approach was also applied to job applicants, who were assessed to determine their suitability for specific careers, often focusing on aircraft pilots. The order of testing varied across participants.

Data Analysis

Data are accessible at Open Science Framework (OSF): : https://osf.io/6m93t/?view_only=cab043ce0c5d4a4fa0a06238ccfc8943. They were analyzed in four steps. In the first step, we looked at mean scores (and standard deviations) on our measures as well as number (and percentage) of individuals in the full sample ($N = 436$) exceeding their cutoffs. We also calculated Pearson product-moment correlations between measures.

Next, we excluded all participants who scored > 9 on the SRSI pseudosymptoms scale (*Yy*) of the SRSI, as scores above this threshold indicate symptom overreporting in psychodiagnostic assessments. We also excluded participants who scored > 22 on the total SS scale because scores beyond that cut point may indicate underreporting. As said before, we further excluded participants based on usual MMPI-2(-RF) cutoffs: CNS > 30, VRIN-r, TRIN-r, and L-r ≥ 80, and F-r ≥ 100 (Aparcero et al., 2022). In this manner, we aimed to ensure that item endorsements on the SRSI genuine symptoms scale in the psychotherapy group were not influenced by overreporting or inattentive/random responses. Similarly, we sought to confirm that item rejections on this scale in the job applicants' group were not driven by underreporting or inattentive/random responses.

In the second step, we only included screened participants (i.e., those with non-deviant scores on validity parameters) and performed a one-way ANOVA, with age as a covariate, on genuine symptoms scores of psychotherapy and job applicants, expecting that the former group would endorse substantially more genuine symptoms on the SRSI that the latter group.

Third, using ROC analyses, we explored for the screened sample optimal cutoffs for the genuine symptoms main scale of the SRSI. The assumption underlying this analysis was that the psychotherapy group had clinically elevated symptom levels that required treatment, whereas job applicants would generally be less symptomatic—i.e., not symptomatic to the extent that would warrant clinical attention. We deemed an AUC of > .90 to be promising enough to justify further research into the diagnostic information conveyed by the genuine symptoms scale of the SRSI.

Fourth, and as an exploratory analysis in the screened sample, we examined Pearson product-moment correlations between the genuine symptoms' subscales and main scale of the SRSI (i.e., *Xx* and its five subscales) and the six MMPI-2-RF scales (i.e., EID and the five somatic/cognitive specific problem scales). In doing so, we applied Bonferroni corrections ($p < 0.05/\ 36 = .0014$). We anticipated that *Xx*, *XDe*, and *XAn* would correlate positively with EID because of its psychopathological content. We also expected that *Xx*, *XCo*, *XPa*, and *XSo* would correlate positively with the five somatic/cognitive specific problem scales. The full correlation matrix is available in the supplemental material.

Results

Full Sample

Table 1 shows mean (*SDs*) scores on our measures as well as Pearson product-moment correlations between them. Regarding these correlations, the observed pattern aligns with expectations: overreporting measures (*Yy* and F-r)[2] show a significant positive correlation with each other, as do the indices of underreporting (SS and L-r). In contrast, overreporting (*Yy* and F-r) and underreporting (SS and L-r) are significantly negatively correlated. Notably, underreporting (i.e., SS) significantly suppresses inattentive responding, as indicated by VRIN-r and TRIN-r. This makes sense, given that individuals who engage in underreporting typically aim to appear highly cooperative.

The percentage of invalid responses—those scoring above the established cutoffs—ranged from 0% (CNS and VRIN-r) to 48% (SS). When applying all cutoffs jointly to exclude invalid responses, a total of 255 participants (58%) were removed, including 45 from the psychotherapy subgroup (31%) and 210 from the job applicants' subgroup

[2] Including other overreporting indices such as FBS-r, Fs, and Fp-r in the exclusion of only one additional participant from the patient group.

Table 1 Mean scores of patients ($n=145$), applicants ($n=291$), and full sample ($N=436$) on measures, numbers scoring above cut points, and correlations between measures

	M (*SD*)	> Cut point *n* (%) patients ($n=145$)	> Cut point *n* (%) Applicants ($n=291$)	> Cut point *n* (%) Full sample ($N=436$)	*Xx Yy*	SS
SRSI *Xx* *Yy* (> 9)	8.07 (9.99) 1.56 (3.88)	21 (14%)	0 (0%)	21 (5%)	.75**	
SS (> 22)	20.58 (6.49)	5 (3%)	205 (70%)	210 (48%)	−.73** −.47**	
MMPI-2-RF	1.20 (2.79)	0 (0%)	0 (0%)	0 (0%)	−.14** −.06	.12*
CNS (> 30)	45.29 (10.22)	0 (0%)	0 (0%)	0 (0%)	.46** .29**	−.47**
VRIN-r (≥ 80)	59.64 (10.59)	20 (14%)	2 (1%)	22 (5%)	.47** .31**	−.46**
TRIN-r (≥ 80)	58.79 (12.23)	0 (0%)	28 (10%)	28 (6%)	−.36** −.22**	.46**
L-r (≥ 80)	50.44 (14.18)	6 (4%)	0 (0%)	6 (1%)	.77** .67**	−.64**
F-r (≥ 100) Combined		24 (17%)	30 (10%)	54 (12%)		
All instruments		45 (31%)	210 (72%)	255 (58%)		

M mean, *SD* standard deviation, *Xx* genuine symptoms scale of the Symptom Self-Report Inventory (SRSI), *Yy* pseudosymptom scale of SRSI, *SS* Supernormality Scale, *CNS* cannot say-responses, *VRIN-r* variable response inconsistencies, *TRIN-r* true response inconsistencies, *L-r* uncommon virtues, *F-r* infrequent responses.

*$p<.05$. **$p<.01$

Table 2 Range, mean and SD of genuine symptoms subscales and main scale scores in both groups

	Patients ($n=100$)			Applicants ($n=81$)		
	Range	*M*	*SD*	Range	*M*	*SD*
Total genuine symptoms (*Xx*)	0–39	16.80	8.32	0–16	2.95	3.02
Cognitive symptoms (*XCo*)	0–9	3.24	2.96	0–4	0.26	0.76
Depressive symptoms (*XDe*)	0–8	2.75	1.88	0–3	0.40	0.70
Pain symptoms (*XPa*)	0–10	2.27	2.24	0–3	0.59	0.74
Unspecific somatic symptoms (*XSo*)	0–10	5.49	2.95	0–6	0.60	1.26
Anxiety symptoms (*XAn*)	0–10	3.05	2.39	0–6	1.10	0.98

M mean, *SD* standard deviation

(72%), resulting in a final screened sample of 181 participants (100 psychotherapy patients, 81 job applicants). Significantly more job applicants failed on the validity indices than psychotherapy patients, Fisher's exact $p<0.01$. Invalid responses on the SRSI pseudosymptom scale (*Yy*) were more prevalent among psychotherapy patients than job applicants (21 vs. 0[3]), whereas the reverse pattern was observed for invalid responses on the SS scale (5 vs. 205). Similar proportions of psychotherapy patients and job applicants failed the MMPI-2-RF validity indices (24 vs. 30,[4] Fisher's exact $p=.07$).

Screened Sample

Table 2 shows the scores (range, mean, and *SD*) of both groups on the genuine symptoms (*Xx*) main and subscales of the SRSI. A series of one-way ANOVAs with age as covariate indicated that psychotherapy patients scored significantly higher on all genuine symptoms' parameters of the SRSI: all Fs (2, 178) > 21.95, all ps < .001. Effect sizes (η^2) for subscale scores ranged from .25 (anxiety symptoms) to .53 (unspecific somatic symptoms). For the main scale, η^2 was .53, with $F(2, 178)=101.13$, $p<.001$.

Classification Accuracy of SRSI Genuine Symptoms

We performed ROC analyses using membership in the psychotherapy group as the positive standard for the main scale and subscales. The AUCs for the subscales were generally below .90, with the unspecific somatic symptoms subscale as an exception (AUC = 0.92, *SD* = .03, 95% CI [.87, .96]). The AUC for the total genuine symptoms scale was satisfactory

[3] We did not calculate Fisher's exact *p*-value here and below due to low frequencies in certain cells.

[4] The combined frequencies of invalid responses on the validity indices exceed the total of 255 participants excluded from the sample, as responses on validity measures may overlap (e.g., an individual might fail both the SS and the L-r indices).

at 0.94 (*SD* = .02, 95% CI [.89, .97]). Table 3 shows various cut points and their corresponding sensitivities, false positive probabilities, and likelihood ratios. For example, a score above 11 on the SRSI *Xx* yields a sensitivity of .75, meaning that at least 75% of individuals with genuine symptoms are correctly identified using this cut point. The false positive rate is .04, indicating that at most 4% of individuals without genuine symptoms are incorrectly classified as having them. The corresponding positive likelihood ratio (+LR) is 18.8, which means that a score above this cutoff is approximately 19 times more likely to occur in someone with genuine symptoms than in someone without them (Rassin et al., 2022).

SRSI Genuine Symptoms and MMPI-2-RF Scales

As expected, psychotherapy patients had higher scores on the EID scale of the MMPI-2-RF than job applicants, means being 18.37 (*SD* = 8.15) and 6.20 (*SD* = 3.97), respectively, $t(179) = 12.30$, $p < .01$, Cohen's $d = 1.89$. *Xx*, *XDe*, and *XAn* were positively correlated with EID ($r = .78$, $r = .74$, and $r = .54$, $ps < .001$). Interestingly, *XSo* also correlated robustly with EID ($r = .78$, $p < .001$). And, as anticipated, *Xx*, *XCo*, *XPa*, and XSo showed moderate to high correlations with their respective MMPI-2-RF-specific problem scale counterparts (see Supplemental table).

Discussion

The main findings of our study can be summarized as follows. First, in the full (i.e., unscreened) sample of psychotherapy participants and job applicants, the SRSI main scales demonstrated findings as expected. Previous research, though based on a small sample ($N = 50$), identified a strong correlation ($r = .81$) between the SRSI pseudosymptoms scale and the F-r scale of the MMPI-RF (in Merten et al., 2022). Within a much larger and more diverse sample ($N = 436$), we replicated this significant correlation, albeit that its size was in the moderate range ($r = .67$). Furthermore, we observed negative correlations between the SRSI main scales and measures of underreporting, specifically the Supernormality Scale (Cima et al., 2003) and the L-r scale of the MMPI-RF (Ben-Porath & Tellegen, 2008/2011).

Second, in screening for overreporting, underreporting, and inattentive/random responding, we excluded more than half of the full sample ($n = 255$; 58%). Distorted responses were particularly prevalent among job applicants. Although we anticipated some degree of underreporting in this group based on previous research (e.g., Butcher et al., 1997), we had not expected underreporting to be so widespread. In hindsight, we have to admit that there exists literature that found similar proportions of distorted responses in job applicants (e.g., Griffith et al., 2005). Evidently, supernormality (e.g., denial of everyday symptoms) is highly common in this group. We suspect that the SS excluded more respondents than the L-r of the MMPI-2-RF because the MMPI has historically been more effective at detecting overreporting than underreporting—especially when the latter is subtle or sophisticated, as operationalized by the SS scale. Indeed, in their meta-analytic study on underreporting indices of the MMPI-2(-RF), Picard et al. (2023; p. 556) concluded: "Research on the detection of exaggerated or feigned psychiatric symptoms using the MMPI-2 and MMPI-2-RF typically generates substantially larger effect sizes (Cohen *d*'s > 2.00; e.g., Rogers et al., 2003; Sharf et al., 2017; Aparcero et al., 2022) for peer review], 2022) than those observed in this meta-analysis ($g = 0.53$ to 1.47). The weaker effect sizes seen in this study indicate that minimizing symptoms is more difficult to differentiate from genuine responding." The only measure within the MMPI framework comparable to the SS is the MMPI-2 S scale, which unfortunately lacks a direct equivalent in the MMPI-2-RF. Taken together, these findings highlight the critical need for measures focusing on symptom underreporting and research concerned with specific norms for psychopathology assessments within clinical populations.

Third, and most importantly, our screened sample provided clear evidence that inspecting the genuine symptoms of the SRSI can yield clinically relevant information. Psychotherapy participants scored significantly higher on both the main scale and subscales of the SRSI genuine symptoms than job applicants, to such an extent that the SRSI genuine symptoms main scale may serve as a tool for identifying cases in need of clinical attention. The AUC (.94) for this main scale was satisfactory, and our exploratory analysis indicated that a cutoff score of 11 can be safely used as a rule of thumb for identifying psychopathology (sensitivity = 75%, false positives < 5%).

Fourth, our findings regarding the relationship between genuine symptoms subscales of the SRSI and MMPI-2-RF

Table 3 Cutoffs for the genuine symptoms main scale of the SRSI and their corresponding sensitivities, false positives, and likelihood ratios

Xx	Cutoff	SENS	FP	+LR
	> 5	.88	.10	8.8
	> 6	.85	.07	12.1
	> 7	.84	.06	14.0
	> 8	.82	.06	13.6
	> 9	.77	.05	15.4
	> 10	.76	.05	15.2
	> 11	.75	.04	18.8

Xx genuine symptoms main scale of the SRSI, *SENS* sensitivity, *FP* false positives (1-specificity), +*LR*, likelihood ratio

indicators should be considered preliminary. Nevertheless, our analyses suggest that *XDe* and *XAn* demonstrated the strongest convergence with MMPI-2-RF psychopathology indicators, while *XCo*, *XPa*, and *XSo* showed their highest convergence with the MMPI-2-RF somatic/cognitive symptom scales. Further replication is needed to confirm the interpretative value of the subscales, but initial findings are promising. In any case, they support the idea that the genuine symptom items of the SRSI capture a broad spectrum of psychopathology.

The positive correlation between the genuine symptoms main scale of the SRSI and the higher-order EID scale of the MMPI-2-RF further supports this conclusion. In summary, our findings suggest that the SRSI genuine symptoms scale offers more than just filler items; it can provide valuable clinical insights, particularly in cases where an individual scores below the cutoff on the pseudosymptom main scale, yet there are concerns about elevated psychopathology levels. This added clinical potential may increase the appeal of the SRSI for clinicians who are hesitant to use symptom validity tests, as such tests are often lengthy, time-consuming, and provide little information when an examinee's pseudosymptom scores are within the normal range (see McCarter et al., 2009; Uiterwijk et al., 2021). However, a caveat is warranted. Our results indicate that the subscales of the SRSI genuine symptoms scale are less effective at screening for elevated levels of psychopathology, as their AUCs were generally low. Naturally, future studies with a broader range of patient and comparison groups are needed to further explore these findings. For the time being, it appears that only the SRSI genuine symptoms main scale is sufficiently robust to serve as a reliable screening tool for psychopathology.

There are several limitations in the current study that merit discussion. First, our sample consisted of French- and Dutch-speaking individuals and included a comparison group that may not fully represent the general population. As a result, the cut points, sensitivities, and false positive rates identified should be considered provisional and interpreted with caution; they primarily highlight that the main scale of SRSI genuine symptoms shows promise and warrants further research.

Second, although both groups were screened using multiple instruments for under- and overreporting, we acknowledge that our screening process was not foolproof. Admittedly, we could have applied additional MMPI-2(-RF) validity indicators with more stringent cutoffs. Thus, we cannot entirely rule out the possibility that even after screening, clinical participants were more inclined to report symptoms, whereas job applicants may have been more motivated to underreport them—potentially inflating the AUC. Nevertheless, our study represents a step in the right direction, as healthy comparison groups in this research area are often unscreened and recruited based solely on self-reported absence of symptoms. Future studies should consider exploring alternative and more rigorous methods for screening comparison groups.

Third, in rigorously excluding individuals who engaged in underreporting, we may have created an artificial "non-symptomatic" group. Certain forms of underreporting—such as lack of insight, fear of stigma, or grandiose denial of symptoms—can themselves be related to psychopathology (e.g., De Page & Merckelbach, 2021). It raises the question of whether individuals with these characteristics should be excluded from samples used to evaluate the effectiveness of a psychopathology instrument. This issue also touches on the broader question of whether overreporting and underreporting should necessarily be viewed as dichotomous phenomena. More articulated conceptualizations of the factors influencing symptom reporting are arguably needed (e.g., Boskovic et al., 2024), and a series of in-depth case studies could contribute significantly to this understanding.

In summary, given the strengths of its genuine symptoms scale, the SRSI holds promise for evolving from a stand-alone symptom validity test into a comprehensive instrument, offering both reliable scales for detecting overreporting and validated measures of psychopathology. Many well-known clinical assessment tools (e.g., MMPI) include validity scales, and tools that lack strong validity scales face limitations in empirical research. The reverse is also true (Ganellen, 2013). While the SRSI is not designed to be a dedicated psychopathology instrument, embedding validated psychopathology scales alongside SVT components strengthens its utility. Future research should focus on replicating and further exploring the interpretative value of its subscales and examining the factor structure of genuine symptom items to reinforce their validity.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12207-025-09543-w.

Data Availability The data associated with this paper are available via the Open Science Framework and can be accessed at [anonymous view-only link for peer review]: https://osf.io/6m93t/?view_only=cab043ce0c5d4a4fa0a06238ccfc8943

Declarations

Ethics Approval The project was approved by the standing ethical committee of the XX University XX (XXX Ref 2020–527).

Conflict of interest H.M. is one of the authors of the original German version of the SRSI, which is commercially available. H.M. and B.D.F. are among the authors of the Dutch version of the SRSI, which is also commercially available.

References

Aparcero, M., Picard, E. H., Nijdam-Jones, A., & Rosenfeld, B. (2022). Comparing the ability of MMPI-2 and MMPI-2-RF validity scales to detect feigning: A meta-analysis. *Assessment, 30*(3), 744–760. https://doi.org/10.1177/10731911211067535

Aryal, K., Merten, T., Akehurst, L., & Boskovic, I. (2022). The English-language version of the Self-Report Symptom Inventory: A pilot analogue study with feigned head injury sequelae. *Applied Neuropsychology: Adult, 31*(5), 1078–1082. https://doi.org/10.1080/23279095.2022.2109158

Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *MMPI-2-RF: User's guide for reports* (2nd ed.); *MMPI-2-RF: Manual for administration and scoring* (rev. ed.). University of Minnesota Press.

Boskovic, I., Giromini, L., Katsouri, A., Tsvetanova, E., Fonse, J., & Merckelbach, H. (2024). The spectrum of response bias in trauma reports: Overreporting, underreporting, and mixed presentation. *Psychological Injury and Law, 17*, 117–128.

Burchett, D., & Bagby, R. M. (2022). Assessing negative response bias: A review of the noncredible overreporting scales of the MMPI-2-RF and MMPI-3. *Psychological Injury and Law, 15*, 22–36.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). MMPI-2: Minnesota multiphasic personality inventory-2: Manual for administration and scoring. University of Minnesota Press.

Butcher, J. N., Morfitt, R. C., Rouse, S. V., & Holden, R. R. (1997). Reducing MMPI-2 defensiveness: The effect of specialized instructions on retest validity in a job applicant sample. *Journal of Personality Assessment, 68*(2), 385–401.

Butcher, J., Derksen, J., Sloore, H., & Sirigatti, S. (2003). Objective personality assessment of people in diverse cultures: European adaptations of the MMPI-2. *Behaviour Research and Therapy, 41*(7), 819–840.

Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., & Muris, P. (2003). The other side of malingering: Supernormality. *The Clinical Neuropsychologist, 17*(2), 235–243.

Czornik, M., Merten, T., & Lehrner, J. (2021). Symptom and performance validation in patients with subjective cognitive decline and mild cognitive impairment. *Applied Neuropsychology: Adult, 28*(3), 269–281.

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.

Dandachi-FitzGerald, B., De Page, L., & Merckelbach, H. (2023a). Detecting symptom exaggeration in compensation-seeking individuals, psychotherapy clients, and individuals referred for job assessments: Psychometric features of the French and Dutch versions of the Self-Report Symptom Inventory. *Psychology & Neuroscience, 16*(3), 261–270.

Dandachi-FitzGerald, B., Pienkohs, S., Merten, T., & Merckelbach, H. (2023b). Detecting symptom overreporting–equivalence of the Dutch and German self-report symptom inventory. *Psychological Test Adaptation and Development, 4*, 85–96.

Dandachi-FitzGerald, B., Merckelbach, H., & Merten, T. (2024). *SRSI Dutch manual*. Hogrefe.

De Page, L., & Merckelbach, H. (2021). Associations between supernormality ('faking good'), narcissism and depression: An exploratory study in a clinical sample. *Clinical Psychology & Psychotherapy, 28*(1), 182–188.

De Page, L., Hennequin, E., & Merckelbach, H. (2013). *French Translation of the Supernormality Scale (SS)*. Internal lab report.

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology, 15*(3), 453–468.

Ganellen, R. J. (2013). *Integrating the Rorschach and the MMPI-2 in personality assessment*. Routledge.

Geurten, M., Meulemans, T., & Seron, X. (2018). Detecting overreporting of symptoms: The French version of the Self-Report Symptom Inventory. *The Clinical Neuropsychologist, 32*(Suppl. 1), 164–181.

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.

Giromini, L., Pasqualini, S., Corgiat Loia, A., Pignolo, C., Di Girolamo, M., & Zennaro, A. (2022a). A survey of practices and beliefs of Italian psychologists regarding malingering and symptom validity assessment. *Psychological Injury and Law, 15*(2), 128–140.

Giromini, L., Young, G., & Sellbom, M. (2022b). Assessing negative response bias using self-report measures: New articles, new issues. *Psychological Injury and Law, 15*(1), 1–21.

Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: An interpretive manual* – (3rd ed.). Allyn & Bacon.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2005). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341–355. https://doi.org/10.1108/00483480710731310

Haber, J. C., & Baum, L. J. (2014). Minnesota multiphasic personality inventory-2 restructured form (MMPI-2-RF) scales as predictors of psychiatric diagnoses. *South African Journal of Psychology, 44*(4), 439–453.

Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2- Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist, 30*(4), 473–496. https://doi.org/10.1080/13854046.2016.1187769

Kirk, J. W., Baker, D. A., Kirk, J. J., & MacAllister, W. S. (2020). A review of performance and symptom validity testing with pediatric populations. *Applied Neuropsychology: Child, 9*(4), 292–306.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 625–630.

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist, 32*(3), 391–421.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2024). Neuropsychological validity assessment beliefs and practices: A survey of north American neuropsychologists and validity assessment experts. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acae102

McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort testing in contemporary UK neuropsychological practice. *The Clinical Neuropsychologist, 23*(6), 1050–1066.

Menton, W. H. (2022). Development and initial validation of differential diagnostic indices for the MMPI-2-RF. *Assessment, 29*(3), 410–424.

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*, 102–111.

Merten, T., Kaminski, A., & Pfeiffer, W. (2020). Prevalence of overreporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients. *The Clinical Neuropsychologist, 34*(5), 1004–1024.

Merten, T., Giger, P., Merckelbach, H., & Stevens, A. (2019). *Self-Report Symptom Inventory (SRSI) – Deutsche Version. Manual*

[German version of the Self-Report Symptom Inventory. Manual]. Hogrefe.

Merten, T., Dandachi-FitzGerald, B., Boskovic, I., Puente-López, E., & Merckelbach, H. (2022). The self-report symptom inventory. *Psychological Injury and Law, 15*, 94–103.

Rassin, E., Arbiyah, N., Boskovic, I., Otgaar, H., & Merckelbach, H. (2022). Likelihood ratios in psychological expert opinion, and their reception by professional judges. *The International Journal of Evidence and proof, 26*(4), 325–341.

Ribatti, R. M., Merten, T., Lanciano, T., & Curci, A. (2024). Psychometric validation of the Italian self-report symptoms inventory (SRSI): Factor structure, construct validity, and diagnostic accuracy. *Applied Neuropsychology: Adult*, 1–16. https://doi.org/10.1080/23279095.2024.2400681

Ribatti, R. M., Merten, T., Lanciano, T., & Curci, A. (2025). Detecting simulated symptoms in chronic pain and fibromyalgia: A two-study evaluation of the Italian version of the Self-Report Symptom Inventory. *Applied Neuropsychology: Adult*, 1–9. https://doi.org/10.1080/23279095.2025.2475951.

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*(2), 160–177. https://doi.org/10.1177/1073191103010002007

Romero, I. E., Toorabally, N., Burchett, D., Tarescavage, A. M., & Glassmire, D. M. (2017). Mapping the MMPI–2–RF substantive scales onto internalizing, externalizing, and thought dysfunction dimensions in a forensic inpatient setting. *Journal of Personality Assessment, 99*(4), 351–362.

Sellbom, M., Bagby, R. M., Kushner, S., Quilty, L. C., & Ayearst, L. E. (2012). Diagnostic construct validity of MMPI-2 restructured form (MMPI-2-RF) scale scores. *Assessment, 19*(2), 176–186.

Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment, 39*, 441–455.

Sleep, C. E., Petty, J. A., & Wygant, D. B. (2015). Framing the results: Assessment of response bias through select self-report measures in psychological injury evaluations. *Psychological Injury and Law, 8*, 27–39.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*, 183–189.

Uiterwijk, D., Wong, D., Stargatt, R., & Crowe, S. F. (2021). Performance and symptom validity testing in neuropsychological assessments in Australia: A survey of practices and beliefs. *Australian Psychologist, 56*(5), 355–371.

Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law, 8*(3), 188–199.