Initial Validation of Short Forms of the SIMS for Neuropsychological Evaluations

Robert J. Spencer [1,2] · Brian D. Gradwohl [1,2] · Victoria M. Kordovski [1]

Abstract

The Structured Inventory of Malingered Symptomatology (SIMS) is a measure of impression management that covers a range of symptomatology through 75 items and five scales, some of which are particularly relevant for cognitive evaluations. The aims of the current projects were to (1) reduce the length of the SIMS to the items most predictive of symptom validity on the Minnesota Multiphasic Inventory-2-RF, (2) explore the factor structure of the remaining items, and (3) test the degree to which the final scale identifies individuals with inadequate performance validity. The sample consisted of 249 veterans referred for outpatient neuropsychological assessment and completed the SIMS, most of whom also completed the MMPI-2-RF and measures of performance validity. Results identified items from the Neurological Injury (11 items) and Amnestic Disorders scales (8 items) that were most strongly related to its theoretically consistent MMPI-2-RF scales and that two distinct factors corresponding to the original scales were identified on factor analysis. The revised subscales correlated at $r = 0.55$, and each scale had acceptable internal consistency ($\alpha = 0.76$–0.84), and the Amnestic and Total Revised SIMS were adequately able to identify individuals failing measures of performance validity. The reliability and validity of the SIMS for Neuropsychological Settings (SIMS-NS) and its Amnestic subscale, despite its brevity, were practically psychometrically equivalent to the original SIMS and one prior short form of the test.

Keywords Malingering · Performance validity · Symptom validity · Assessment · Structured inventory of malingered symptomatology

As a collaborative enterprise between examiner and examinee, the utility of clinical neuropsychological assessments depends largely on obtaining functionally relevant data through observing performances on standardized tests and interpreting self-report instruments. In clinical contexts, patients often complete structured checklists requiring self-reflection on aspects of symptom burden, emotional status, and functional capacity (Rabin, Paolillo, & Barr, 2016). These self-report instruments can provide valuable data regarding symptoms that may not be apparent during structured, performance-based assessments. Furthermore, because examinees are completing these measures with reference to their everyday lives, scores may reflect overall changes in symptomatology and its impact on daily living in an ecologically relevant manner. The scores generated from such instruments are objective and can be compared to reference groups and within the same individuals across multiple assessments.

Despite clear advantages, neuropsychologists interpreting structured checklists must bear in mind that data are influenced by other factors, such as emotional burden, attentiveness to the content of items, and, most relevantly to the current project, the veracity of self-reports. Broadly, the representativeness of reporting style on these instruments is referred to as symptom validity (SV). In the neuropsychological context, SV usually pertains to the degree to which examinees provide credible accounts of their symptomatology, devoid of exaggeration or fabrication (Bianchini, Mathias, & Greve, 2001). Several instruments have been devised to detect violations of SV, and such measures are referred to as symptom validity tests (SVT). Similarly, performance validity tests (PVT) refer to measures that purportedly identify individuals not responding to laboratory-based tests with credible performances that match their true capabilities. SVT and PVT, while

✉ Robert J. Spencer
  Robert.Spencer2@va.gov

1 Mental Health Service, VA Ann Arbor Healthcare System, 2215 Fuller Road, Ann Arbor, MI 48105, USA

2 Department of Psychiatry, University of Michigan Health System, Ann Arbor, MI 48105, USA

conceptually similar, often correlate only modestly with one another, suggesting that they represent distinct but overlapping aspects of protocol validity (Larrabee, 2012).

Among the most frequently used SVTs is the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005). Initially developed by Smith and Burger in 1997, the SIMS was created as a screening instrument to detect feigned psychopathology and consists of five categories (i.e., subscales) that were believed to be among the most commonly feigned disorders. Items were chosen based on the knowledge of conditions at the time of its development, useful items from other instruments sensitive to malingering (e.g., Minnesota Multiphasic Personality Inventory), and opinions from a panel of clinical psychologists. The final measure contains 75 true/false items covering the five categories, including Neurologic Injury, Affective Disorders, Psychosis, Low Intelligence, and Amnestic Disorders. Items for each category appear at some level to represent a plausible symptom but are infrequently endorsed by individuals with confirmed diagnoses within that category. For example, although some items for the category of Psychosis might appear to be typical of schizophrenia to a significant proportion of examinees, these items are not endorsed frequently by individuals diagnosed with schizophrenia. Through this approach, the authors of the test offer cutoff scores beyond which individuals with confirmed disorders rarely score, but individuals simulating the same disorders score with relatively greater frequency.

The SIMS has good internal consistency (Widows & Smith, 2005) and adequately identifies individuals with problematic response styles (van Impelen, Merckelbach, Jelicic, & Merten, 2014). Some research has found that the test is more useful for flagging potentially exaggerated self-reports than it is for confirming such instances. For example, Mazza et al. (2019) found that the SIMS Total score, Neurologic Impairment scale, and Low Intelligence Scale discriminated between individuals with genuine psychopathology who were deemed by clinical raters to be simulating symptoms, were exaggerating their genuine symptoms, or were representing their symptoms. In the research by Parks, Gfeller, Emmert, and Lammert (2016), the SIMS total score and Affective Disorders scale were able to differentiate feigned PTSD, post-concussive syndrome, and combined PTSD and post-concussive syndrome presentations. Not only is the SIMS reflective of particular types of responding, but some scales may also be more suited to discrete populations.

Although the SIMS offers five scales, not all of the scales are equally relevant across contexts. Some scales are more appropriate for some purposes than others, indicating that users should consider the test scale by scale and not as an immutable whole. For instance, an examiner evaluating potentially feigned psychosis or intellectual disability in a pretrial forensic evaluation might focus disproportionately on the Psychosis and Low Intelligence scales, whereas

Neurological Impairment and Amnestic Disorders scales would be of more interest in a neuropsychology setting. Indeed, the Neurologic Impairment, Amnestic Disorders, and Affect Disorders subscales are particularly sensitive in patients with psychogenic non-epileptic events (Benge et al., 2012) and psychogenic neurologic disorders (van Beilen, Griffioen, Gross, & Leenders, 2009). The development of a tailored measure that could be applied in neuropsychological populations would fill a large gap in the literature of PVTs and neuropsychology, improve efficiency in clinical assessment, and potentially improve detection of selectively feigned cognitive impairment.

Acknowledging the potential advantages of the SIMS, its practical application is limited by its length. Compared to many of the checklists quantifying symptoms such as depression, anxiety, alcohol use, and postconcussive experiences, which typically contain fewer than 30 items, the 75 items of the SIMS constitute a significant time commitment for all involved. From a practical standpoint, one could argue that if some of the items of the SIMS were unnecessary and could be eliminated, the test would be more portable and could be used more widely. Since its development, few attempts have been made to shorten the instrument. Rogers, Robinson, and Gillard (2014) developed two feigning scales, Rare Symptoms (SIMS-RS) and Symptom Combination (SIMS-SC), to truncate the SIMS. The SIMS-RS scale consisted of items endorsed by fewer than 10% of genuine responders but greater than 25% of feigners, whereas the SIMS-SC scale consisted of pairs of items that are infrequently endorsed by genuine responders but were more commonly endorsed in feigners. Although the initial study demonstrated excellent classification statistics, with quite large effects observed on the combination scale, these initial results were not replicated (Edens, Truong, & Otto, 2020). Malcore, Schutte, Dyke, and Axelrod (2015) created a shortened form (SIMS-SF) by removing items that were weakly related to the total SIMS score or were invariably endorsed in a particular direction. The final scale consisted of 35 items and performed similarly to the full SIMS in discriminating over-reporters. This study, too, has yet to be replicated.

This project examined the feasibility of amending the SIMS within a neuropsychological assessment context. We examined the psychometric properties of the SIMS, including its reliability and utility in predicting failures on tests of performance validity. We evaluated each item of each scale according to its relationship with scales from the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2011), which contains the most frequently used measures of symptom validity (Rabin et al., 2016). We will subject items of sufficient correlations with the MMPI-2-RF to the analyses of internal consistency, factor structure, and validity to predict scores on a composite measure of performance validity. Provided the

revised scales have predictive validity, we will subject these revised scales to logistic regression to offer the reader interpretive guidelines. We hypothesized that the revised version of the SIMS will be considerably shorter than the original version, while retaining its reliability and predictive validity. Furthermore, because we are employing a sample of individuals evaluated for neuropsychological assessment, we predicted that items from some scales would demonstrate acceptable psychometric properties, whereas others will not. We will also compare the psychometrics of our new scale, the SIMS-NS, to those from the standard SIMS and prior short forms of the test (SIMS-RS, SIMS-SC, and SIMS-SF).

Method

We examined SIMS data in a sample of 263 veterans who completed an outpatient neuropsychological evaluation at a Midwestern Veterans Affairs Medical Center. Four participants were excluded for missing data (as defined by SIMS manual instructions) and two for not completing the MMPI-2-RF or a performance validity test (PVT), and six cases were excluded due to invalid MMPI-2-RF Variable Response Inconstancy and/or True Response Inconsistency (i.e., scores exceeding 79; Ben-Porath & Tellegen, 2011). The final sample included 249 veterans (88.4% male) with an average age of 52.6 (SD = 15.0) and 13.7 years of education (SD = 2.2). Primary diagnoses included psychiatric/mood disorder (36.9%), mild neurocognitive disorder (14.1%), major neurocognitive disorder (3.6%), developmental disorder (5.6%), sleep disorder (3.2%), other (4.4%), and no diagnosis (32.1%). A subsample of 144 veterans, including two participants who did not complete a PVT, also completed the MMPI-2-RF. The subsample (88.0% male) had an average age of 49.9 (SD = 14.8) with 13.9 years of education (SD = 2.1). Primary diagnoses of the subsample included psychiatric/mood disorder (42.4%), mild neurocognitive disorder (9.7%), developmental disorder (6.3%), sleep disorder (3.5%), other (3.5%), and no diagnosis (34.7%). The study was approved by the hospital IRB.

Measures

The SIMS is a 75 true/false item measure of symptom validity that takes 10–15 min to complete (Widows & Smith, 2005). The SIMS has adequate internal consistency (e.g., $\alpha = .72$, Merckelbach & Smith, 2003). The SIMS contains a total score and five subscales of 15 items each: Neurologic Impairment, Affective Disorders, Psychosis, Low Intelligence, and Amnestic Disorders. Higher scores reflect a greater probability of invalid symptom endorsement. The Rare Symptom (SIMS-RS), Symptom Combination (SIMS-SC), and Malcore and colleagues' short form (SIMS-SF) will also be calculated for

comparisons. The SIMS-RS is a 15-item scale consisting of uncommonly endorsed items (3, 8, 14, 21, 28, 34, 40, 42, 56, 58, 63, 67, 69, 73, and 75), and SIMS-SC contains 13 pairs of items in which a point is awarded for a pair only when both responses indicate feigning (4–74, 5–60, 20–24, 25–61, 25–64, 27–61, 35–60, 47–59, 47–74, 48–65, 52–66, 56–67, and 56–75; Rogers et al., 2014). The SIMS-SF is a 35-item scale that was developed by empirically reducing items based on corrected item-total correlation and infrequent responding (1, 5, 6, 12, 13, 15, 17, 18, 19, 20, 22, 25, 26, 27, 29, 30, 31, 33, 35, 36, 37, 39, 45, 47, 49, 50, 51, 53, 59, 60, 61, 62, 64, 66, and 71; Malcore et al., 2015).

The MMPI-2-RF is a 338 true/false item measure of personality and psychopathology that takes 35–50 min to complete (Ben-Porath & Tellegen, 2011). The MMPI-2-RF contains several validity scales used to measure over-reporting of symptoms for general (F-r), psychiatric (Fp-r), medical (Fs), somatic and cognitive (FBS-r), and memory (RBS). Based on consensus from the authors, SIMS subscales and MMPI-2-RF validity scales were matched if conceptually related. See Table 1 for details.

Performance validity tests (PVTs) included the Test of Memory Malingering (TOMM; Tombaugh, 1996), Word Memory Test (WMT; Green, 2003), Dot Counting Test (DCT; Boone, Lu, & Herzberg, 2002), Word Choice Test (WCT; Pearson, 2009), Timed Digit Span (TDS; Babikian, Boone, Lu, & Arnold, 2006), Digit Span Scaled Score (DSSS; Spencer et al., 2013; Wechsler, 2008), and the Effort Index from the Repeatable Battery for the Assessment of Neuropsychological Status (Randolph, 1998, 2012; Silverberg, Wertheimer, & Fichtenberg, 2007). Participants completed a median of 2 unique PVTs ($M = 2.51$, SD = 1.11, range 1–5). The full sample was split into valid ($n = 194$) and invalid ($n = 55$) groups. The invalid performance

Table 1 SIMS Subscale and corresponding MMPI-2-RF Validity Scales

SIMS	MMPI-2-RF
NI	F, FBS, Fs
AF	F, Fp-r
P	F, Fp-r
LI	F
AM	F, RBS

SIMS Structured Inventory of Malingered Symptomatology, *NI* Neurological Impairment, *AF* Affective Disorders, *P* Psychosis, *LI* Low Intelligence, *AM* Amnestic Disorders, *MMPI-2-RF* Minnesota Multiphasic Personality Inventory-2 Restructured Form, *F* Infrequent Responses, *FBS* Symptom Validity Scale, *Fs* Infrequent Somatic Responses, *Fp-r* Infrequent Psychopathology Responses

40

Psychol. Inj. and Law (2021) 14:37–45

was defined by "standard" failure on more than one PVT or an "egregious" failure on any PVT. Standard failures were based primarily on the respective manuals and/or established cutoffs in the empirical literature: ≤ 40 for TOMM Trial 1 or ≤ 45 on Trial 2 or Retention, $\leq 82.5\%$ on WMT Immediate Recall or Delayed Recall or $\leq 75.0\%$ Consistency, ≥ 14 on DCT, ≤ 41 on WCT, ≥ 3.5 on TDS, ≤ 4 on DSSS, and > 3 on RBANS-EI. Egregious failure was set at ≤ 30 on any TOMM trial, $\leq 60.0\%$ on WMT primary trials, ≥ 22 on DCT, ≤ 30 on WCT, ≥ 5.5 on TDS, ≤ 2 on DSSS, and > 6 on RBANS-EI. In our clinical setting, each individual failing one PVT, unless egregious, was administered additional PVTs. Thus, those failing only one PVT were categorized as being in the valid group.

Data Analysis

We began by examining data for the standard SIMS in the full sample, with analyses including descriptive statistics, internal consistency, and the proportion of individuals who, according to the total score, would be flagged as giving invalid self-reports. We also examined the demographic differences using chi-squared test and analysis of variance between the valid and invalid groups. The item reduction phase began in the subsample of 144 individuals who completed the MMPI-2-RF. Each item was evaluated based on its point-biserial correlation with its relevant MMPI-2-RF scale(s) based on conceptual relatedness, as agreed upon by the authors. Given that F-r represents broad over-reporting, all items on the SIMS were compared with F-r. Additionally, Neurological Injury items were compared with FBS-r and Fs, since both of the latter scales measure infrequent cognitive and/or somatic complaints that would plausibly be over-reported in feigned neurological disorders. Affective Disorders items were compared with FP-r, and the Psychosis items were compared with FP-r, as this scale captures infrequent symptoms in psychiatric populations. Finally, Amnestic Disorders items were compared with RBS, because of the scale's attempt at detecting exaggerated memory complaints, specifically. Low Intelligence items were only compared with the Fr, as no scale within the MMPI-2-RF was considered to be directly relevant to symptom validity pertaining to intellectual deficits. Items correlating with at least one relevant MMPI-2-RF scale at 0.40 and above were retained. Additionally, items were retained for correlations exceeding 0.30 on more than one relevant MMPI-2-RF scale. The final item pool was then subjected to factor analysis to determine the underlying factor structure of the new scale.

The new scale and the original scale are compared with respect to internal consistency and the validity of each to predict failures on measures of performance validity. These analyses were also applied to other truncated forms, SIMS-RS, SIMS-SC, and SIMS-SF, for comparisons. The "symptoms" most relevant in neuropsychological settings tend to pertain to

cognitive problems. Accordingly, to provide the most useful predictive utility of the SIMS score as a screening measure in a neuropsychological context, performance on tests of performance validity, as opposed to traditional symptom validity, was considered the primary criterion variable. Aside from the MMPI-2-RF, no other SVT was available for cross validation, and re-using the MMPI-2-RF in this sample for validation of the new SIMS scales would risk inflating the resulting validity coefficients. Although conceptually distinct, PVTs served as an independent criterion of protocol validity against which to evaluate the relative accuracy of the various SIMS iterations. Performance validity failure was defined as failing more than one standard PVT or committing at least one egregious failure. An adequate predictive utility was defined as having a receiver operating characteristic curve exceeding 0.70 (Hosmer & Lemeshow, 2000). Finally, logistic regression with base rate transformations will be presented to aid in the interpretation of the new scale. In the logistic regression analyses, probabilities are calculated based on only the scores for each SIMS variation, and then these resulting probabilities are algebraically transformed according to three sufficiently distinct base rates (i.e., 10%, 30%, and 50%). These values, which reference specific scores, are interpreted in a similar fashion as positive predictive power, which uses groupings.

Results

The valid and invalid groups did not have any significant demographic differences in terms of age, $F(1, 248) = 3.32$, $p = .07$, years of education, $F(1, 248) = 2.25$, $p = .14$, and sex, $X^2(1, 249) = 1.3$, $p = .25$. Table 2 presents additional descriptive statistics for the sample, SIMS, and the MMPI-2-RF. Total SIMS scores ranged from 4 to 52, and the mean score for the total score and each of the 5 components scales were either at or above the cutoff score recommended in the manual. By contrast, the mean scores from each of the MMPI-2-RF scales were below the point at which the MMPI-2-RF manual suggests is a problematic level. Regardless of the cutoff score used, the rate of "positive" scores was relatively high for the Total SIMS, with 58.2%, 48.6%, 35.7%, and 21.7% of individuals exceeding the cut scores of > 14, > 16, > 19, and > 23, respectively.

Table 3 presents the intercorrelations between scales of the SIMS. Overall, the scales correlated with each other between .24 and .54, with the lowest correlation observed between Low Intelligence and Amnestic Disorders and the strongest correlation occurring between Amnestic Disorders and Neurologic Impairment. The internal consistencies for each scale, measured by Cronbach alpha, were .86 (total), .77 (NI), .37 (AF), .69 (P), .48 (LI), and .80 (AM).

A subsample ($n = 144$) of participants completed SIMS and MMPI-2-RF. There was a significant age difference,

Table 2 Descriptive statistics

	Mean	SD	Range
Age	52.6	15.0	23–87
Education	13.7	2.2	8–20
SIMS Total	17.7	8.8	4–52
NI	4.0	3.1	0–15
AF	5.9	2.5	1–12
P	1.0	1.6	0–9
LI	2.2	1.8	0–10
AM	4.6	3.3	0–14
MMPI-2-RF[a]			
VRIN-r	53.4	10.4	34–77
TRIN-r	57.4	6.3	50–73
F-r	74.1	21.1	42–120
Fp-r	60.5	17.5	42–420
Fs	69.6	19.5	42–120
FBS-r	68.9	15.1	42–108
RBS	78.0	17.2	42–120
L-r	57.7	10.9	37–95
K-r	44.1	10.1	24–69

$N = 249$

SD standard deviation, *NI* Neurologic Impairment, *AF* Affective Disorders, *P* Psychosis, *LI* Low Intelligence, *AM* Amnestic Disorders

[a] $n = 144$

$F(1,247) = 9.54$, $p = .002$, between those that completed the MMPI-2-RF ($M = 50.04$, SD $= 14.8$) and those that did not ($M = 55.9$, SD $= 14.6$). Completers and non-completers did not differ with respect to years of education, gender, and SIMS total score (ps $> .05$). Table 4 depicts the correlation of SIMS and MMPI-2-RF scales. Generally speaking, no scale had a large correlation with L-r of the MMPI-2-RF or with the Low Intelligence scale of the SIMS. Comparing across all MMPI-2-RF scales pertaining to over-reporting, median

Table 3 SIMS intercorrelations

	NI	AF	P	LI	AM
SIMS Total	.77	.73	.57	.53	.83
NI	-	.39	.26	.28	.54
AF		-	.34	.29	.51
P			-	.32	.36
LI				-	.24
AM					-

$N = 253$; all significant at $p < .001$

SIMS Structured Inventory of Malingered Symptomatology, *NI* Neurologic Impairment, *AF* Affective Disorders, *P* Psychosis, *LI* Low Intelligence, *AM* Amnestic Disorders

Table 4 SIMS and MMPI-2-RF Pearson correlations

	SIMS total	NI	AF	P	LI	AM
F-r	.77	.64	.64	.46	.35	.60
Fp-r	.58	.38	.56	.47	.33	.41
Fs	.60	.63	.47	.23	.23	.44
FBS-r	.58	.51	.54	.31	.26	.39
RBS	.73	.58	.62	.41	.32	.62
L-r	− .18	− .15	− .18	− .08	− .11	− .12
K-r	− .44	− .29	− .51	− .33	− .17	− .31

$n = 144$

correlations for the SIMS include .60 (Total), .58 (Neurological Impairment), .56 (Affective Disorders), .41 (Psychosis), .32 (Low Intelligence), and .44 (Amnestic Disorders). Generally speaking, each SIMS scale was relatively more strongly correlated with its hypothesized MMPI-2-RF validity scales than with other MMPI-2-RF validity scales.

The process of item refinement began with comparing each item on the SIMS with its group of corresponding MMPI-2-RF scales that were hypothesized to be conceptually relevant. Items were retained if they correlated with any conceptually relevant MMPI-2-RF scale at .40 or above or if they correlated with more than one of the conceptually relevant MMPI-RF-scales at .30 or greater. A total of 24 items were identified, with 11 items retained for Neurological Injury, 8 retained for Amnestic Disorders, 3 retained from Affective Disorders, 2 retained from Psychosis, and no items retained from Low Intelligence. Because the goal of the study was to develop conceptually relevant scales, Affective Disorders, Psychosis, and Low Intelligence were eliminated, as there were too few items to constitute separate scales.

To test whether the underlying conceptual structure of the 19 final items matched the hypothesized domains from the original test, all items were subjected to factor analysis. Using a principal component analysis with oblique promax rotation with Kaiser normalization, 5 factors had eigenvalues over 1.0, but examination of the scree plot indicated two stable factors. We therefore specified a two-factor solution, which accounted for 35.4% of the variance. All items clearly loaded on its original factor, with loadings of .40 or greater. The lone exception was a cross loading for item 45, which was retained on its original scale for the sake of consistency.

The final SIMS for Neuropsychological Settings (SIMS-NS) consists of 11 items (1, 5, 26, 35, 39, 50, 59, 64, 66, 71, and 74) for Neurological Impairment (NI-11) and 8 items (22, 25, 27, 30, 33, 36, 45, and 49) for Amnestic Disorders (AM-8). We also computed a total combined score. In every case, items are scored with a "true" response, as none of the 12 items from the SIMS scored in the "false" direction were retained. The internal consistencies of the new scale and its original counterpart are comparable, as illustrated in Table 5.

42

Psychol. Inj. and Law (2021) 14:37–45

Table 5 Internal consistency/Cronbach's alpha

	Original SIMS (item total)	SIMS-NS (item total)	SIMS-RS (item total)	SIMS-SC (item total)	SIMS-SF (item total)
Total	.86 (75 items)	-	.56 (15 items)	.63 (13 pairs)	.86 (35 items)
NI+AM	.86 (30 items)	.84 (19 items)	-	-	-
NI	.77 (15 items)	.76 (11 items)	-	-	-
AM	.80 (15 items)	.78 (8 items)	-	-	-

N = 253

SIMS Structured Inventory of Malingered Symptomatology, *NS* Neuropsychological Settings, *RS* Rare Symptoms, *SC* Symptom Combination, *SF* Short Form, *NI* Neurologic Impairment, *AM* Amnestic Disorders

Internal consistencies for the revised subscales are adequate (α = .76–.84), and the SIMS-NS has an internal consistency of .84, which is adequate. In comparison, the SIMS-SF had similar internal consistency of .86, and the SIMS-RS and SIMS-SC had lower internal consistency coefficients of .56 and .63, respectively.

The original SIMS, SIMS-NS, SIMS-RS, SIMS-SC, and SIMS-SF were subjected to predictive analyses using receiver operating characteristics area under the curve (AUC) analyses. As is apparent on Table 6, the original, SIMS-NS, SIMS-SC, and SIMS-SF serve as adequate screens for failures of performance validity, and the SIMS-RS demonstrated unacceptable classification accuracy. In this sample, there was a 22.1% base rate of invalid performances, and the overall SIMS-NS had an AUC of 0.74 (0.66–0.81). The AM-8 and NI-11 subscales had AUCs of 0.74 (0.66–0.81) and 0.67 (0.59–0.75), respectively.

To place the predictive utility of both the original and revised SIMS scores on continua, each scale was subjected to binary logistic regression so that the probability of performance validity failure could be estimated at each potential score of the two revised versions of the SIMS with the most predictive utility according to AUC analyses. Regardless of the base rate, SIMS-NS, AM-8, and NI-11 scores of 6, 3, and 4, respectively, equate to elevated odds of failing measures of performance validity. However, the absolute probability of performing invalidly is anchored to the estimated base rate of the population to which it is applied. Table 7 presents the relative probabilities of performance validity failure according to base rates of 10%, 30%, and 50%. Depending on the base rate estimation, SIMS for Neuropsychological Settings Total scores ranging from 6 to 17 crests the 50% probability range for invalid performance validity. For the SIMS-NS AM-8 subscale, a score of 3 exceeds the 50% threshold when the base rate is 50%, and a score of 6 exceeds a 50% probability of failing a performance validity test when applied to a population with a 30% base rate of PVT failures. For the SIMS-NS NI-11 subscale, a score of 4 exceeds the 50% threshold when the base rate is 50%, and a score of 7 exceeds a 50% probability of failing a performance validity test when applied to a population with a 30% base rate of PVT failures. By comparison, total scores of 20 or higher on the original SIMS are associated with elevated risk of failing performance validity, and scores of 40, 28, and 20 are needed to exceed 50% probability of performing invalidly at base rates of 10%, 30%, and 50%, respectively.

Discussion

The 19-item SIMS for Neuropsychological Settings has equivalent reliability and predictive validity as the standard SIMS but contains 56 fewer items, and its content is specific to neuropsychological evaluations. The shortened Amnestic subscale (AM-8), which consists of only 8 items, is also a potentially feasible alternative to the lengthier SIMS, its subscales, or even the truncated SIMS-NS. The brevity and comparable psychometrics of the shortened versions of the SIMS makes for more portable alternatives to offer as brief measures of impression management for neuropsychological assessments.

As with the original SIMS, most truncated versions of the SIMS have acceptable internal consistency. The SIMS-NS and the AM-8 subscale are adequate predictors of performance validity, and each is equivalent to the predictive utility of the full instrument and other shortened versions of the SIMS, including the Rare Symptoms scale, Symptom Combination scale, and SIMS-Short Form. The 11-item Neurological Impairment (NI-11) subscale demonstrated marginally weaker predictive ability compared with the AM-8 subscale and full SIMS-NS. Even though performance validity and symptom validity are conceptually distinct, there is some overlap in the constructs and how they are interpreted for individuals, and most iterations of the SIMS serve adequately to flag individuals with potential failures of performance validity tests. As with all measures, predicting the probability of failing measures of performance validity depends on the base rate of the population to which the measures are applied. Using logistic regression and algebraic adjustment for base rate, scores of 17, 11, and 6 for the truncated scale is associated with at least a 50% chance of performing invalidly at base rates of 10%, 30%, and 50%. Because it is advisable to avoid pejorative labels and potential stigma,

Table 6 ROC-AUC analysis for PVT failure

	Original SIMS AUC (95% CI)	SIMS-NS AUC (95% CI)	SIMS-RS AUC (95% CI)	SIMS-SC AUC (95% CI)	SIMS-SF AUC (95% CI)
Total	0.76* (0.69–0.83)	–	0.54 (0.45–0.62)	0.72* (0.63–0.80)	0.76* (0.69–0.83)
NI + AM	0.75* (0.69–0.82)	0.74* (0.66–0.81)	-	-	
NI	0.67* (0.59–0.76)	0.67* (0.59–0.75)	-	-	
AM	0.76* (0.69–0.83)	0.74* (0.66–0.81)	-	-	

$n = 253$; 21.7% base rate of invalid performance validity; $* = p < .001$

SIMS Structured Inventory of Malingered Symptomatology, *NS* Neuropsychological Settings, *RS* Rare Symptoms, *SC* Symptom Combination, *SF* Short Form; *NI* Neurologic Impairment, *AM* Amnestic Disorders

performance standards on tests of protocol validity are usually adopted that maximize positive predictive power and specificity, even at the cost of negative predictive power and sensitivity. For this reason, it is probably best to regard these performance standards as screens for invalid symptom reporting and not to interpret such scores as proof positive of negative impression management. SIMS-NS scores of 16 and 12 equate to a probability of equal to or greater than 75% chance of invalid performance when base rates are 30% and 50%, respectively, and no score equates to 75% PVT failure when a 10% base rate is used. Those using the standard SIMS, SIMS-NS, and its newly developed subscales can reference Table 7 to estimate

the probability of failing a performance validity test. The reader is cautioned, however, that these statistics are best suited to clinical neuropsychological assessments, and results should not be assumed to apply to other clinical and/or forensic contexts until cross-validating research provides sufficient justification for its use.

Regardless of its form, the SIMS is probably regarded best as an instrument for identifying individuals potentially providing invalid responses. Examiners should follow up with additional measures of symptom validity. Regarding performance validity, scores on the SIMS can refine predictions of the relative likelihood of performance validity failure. Some

Table 7 Probability of performance validity failure across various base rates

Score	SIMS-NS base rate percentage			AM-8 base rate percentage			NI-11 base rate percentage		
	10%	30%	50%	10%	30%	50%	10%	30%	50%
0	3	12	23	4	12	25	5	17	32
1	4	14	27	5	17	32	6	21	38
2	5	16	31	7	23	41	8	25	44
3	6	19	36	10	30	*50*	10	30	50
4	7	23	40	14	39	60	12	35	*56*
5	9	26	45	19	48	68	15	41	62
6	10	30	*50*	26	*57*	*76*	19	47	67
7	12	35	55	34	66	82	23	*53*	72
8	14	39	60	43	74	87	27	59	*77*
9	17	44	65				32	65	81
10	20	49	69				38	70	85
11	23	*54*	73				44	*75*	88
12	27	59	*77*						
13	31	64	80						
14	36	68	83						
15	40	72	86						
16	45	*76*	88						
17	*50*	80	90						
18	55	83	92						
19	60	85	93						

Italicized items indicate the 50% and more conservative 75% threshold

NS Neuropsychological Settings, *AM* Amnestic Disorders Subscale, *NI* Neurologic Impairment

studies have found value in combining symptom and performance validity measures to predict breaches and protocol validity (Spencer et al., 2017). Additional research should examine the predictive power of combining SVTs and PVTs with the original and truncated versions of the SIMS.

In addition to having a strong relationship with performance validity, the iterations of the SIMS share strong correlations with the validity scales of the MMPI-2-RF. Interpretively, symptom validity scales should never replace performance validity scales, even when symptom validity scales strongly predict performance validity failure. In this study, as with other studies (Larrabee, 2012), symptom validity and performance validity scales are positively related but imperfectly so. In effect, four categories are created that correspond to passing both types of measures, failing both, failing one, or failing the other. In two of these four categories, validity data are confirming of each other, whereas in other instances, the information is apparently discrepant. In clinical practice, it is quite common to see instances of over-reporting, but with performances that are adequate, and vice versa. Such discrepancies lead to interpretive complexities. Despite the term "malingered" in the SIMS, no motivation should be assumed when interpreting results in isolation. Clinicians interpreting protocol validity should consider measures of performance validity and symptom validity within contextual factors such as incentives, potential psychological gain, and other sociocultural contingencies.

Importantly, the current study adds to the literature on prior attempts at shortened forms of the SIMS. First, the findings address prior research showing modest classification accuracy for the SIMS-RS (Edens et al., 2020; Rogers et al., 2014). Given the SIMS-RS data in this sample, and the lack of overlap with items in the SIMS-NS, our findings do not support the use of SIMS-RS in neuropsychological settings. In contrast, Rogers and colleagues' SIMS-SC scale showed better classification accuracy and was psychometrically comparable to the SIMS-NS. However, the SIMS-SC scale requires effort to score, as it involves comparing 13 pairs of items, and the additional time and effort may deter some clinicians seeking a quick screening measure. Finally, the SIMS-SF, proposed by Malcore et al. (2015), had comparable internal reliability and PV classification accuracy as the full SIMS and the SIMS-NS (see Table 5), and all were more reliable and accurate than the SIMS-RS and SIMS-SC scales. This congruence may be driven by the high item overlap, as 18 of the 19 items on the SIMS-NS are also on the SIMS-SF.

Users of the SIMS-NS scales should be aware of potential consequences of structural changes to the SIMS. The content of items for the SIMS ranges from appearing innocuous to implausible. It is important to remember that the item selection process in this study did not consider the content of items, only the scale to which it belonged and the correlations with other tests. In this way, we adopted a relatively pure empirical keying approach to item refinement.

The SIMS-NS scales cover a narrower range of symptoms than the standard SIMS and are therefore more tailored to cognitive evaluations, compared with broader clinical settings. Even though the total number of items was reduced from 75 to 19, one could argue that simply using the Neurological Injury and Amnestic Disorders scales of the standard SIMS would be sufficient for detecting negative response bias. This is largely true, although the SIMS-NS eliminates 11 of the worst-performing items in the standard subscales. The standard SIMS subscales do not detract from the psychometrics of the SIMS-NS subscales, but neither do they add predictive value.

The shorter scales greatly reduce administration time and burden on examinees, but users should be aware that administering a shortened measure does not circumvent copyright law and that full forms should be purchased for each use, regardless of the length of the scale.

This study, and by extension the SIMS-NS, has several limitations. This sample was diagnostically heterogeneous, and confirmation of the validity of the SIMS-NS is needed across contexts with samples that are more homogeneous. Conversely, the sample was demographically homogeneous, and additional studies with greater ethnoracial and educational diversities are needed. Although this study found that the SIMS adequately detected performance validity failures, a more relevant test for the SIMS and its shorter iterations would be to assess its relationship with symptom validity scales. In this study, we were not able to test this question with the MMPI-2-RF, as the MMPI-2-RF was used to select the SIMS-NS items, and would therefore produce an inflated estimate of its relationship. This issue needs to be assessed in cross-validating research with an independent sample.

Each of the items on the refined scale is keyed in the true direction, which aids in rapid interpretation. Problematically, however, it can be argued that this scoring schema makes it easier for examinees who are exaggerating to surmise the direction of the "pathological" scores and provide responses accordingly. Overall, we observed that the 12 items that were reversed scored tended to have lower item-total correlations with the remainder of the test, potentially indicating that although attractive in principle, reverse-scoring in this context simply muddies the waters. Given the possible impact of the short form on the instrument's face validity, it will be important for future validation studies to examine this new measure in lieu of the full SIMS.

In conclusion, the SIMS-NS has comparable reliability and validity as the full SIMS, albeit with 56 fewer items and a focus on neuropsychological symptomatology. The 8-item Amnestic Disorders scale (AM-8) offers an even shorter alternative, without substantially sacrificing psychometric precision.

Compliance with Ethical Standards

Disclaimer This work was authored as part of the contributor's official duties as an employee of the US Government and is therefore a work of the US Government. In accordance with 17 U. S. C 105, no copyright protection is available for such works under US Law.

Conflict of Interest The authors declare that they have no conflicts of interest.

Ethics Approval The study was approved by the facility's IRB.

Consent to Participate The participants consented to their clinical data being used for research purposes.

Consent for Publication Not applicable.

Code Availability Not Applicable

References

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145–159. https://doi.org/10.1080/13854040590947362.

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: a critical review. *The Clinical Neuropsychologist, 15*(1), 19–45.

Ben-Porath, Y. S., & Tellegen, A. (2011). *The Minnesota multiphasic personality inventory -2 Restructured form: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Benge, J. F., Wisdom, N. M., Collins, R. L., Franks, R., LeMaire, A., & Chen, D. K. (2012). Diagnostic utility of the structured inventory of malingered symptomatology for identifying psychogenic non-epileptic events. *Epilepsy & Behavior, 24*(4), 439–444.

Boone, K., Lu, P., & Herzberg, D. (2002). *The dot counting test*. WPS publishing.

Edens, J. F., Truong, T. N., & Otto, R. K. (2020). Classification accuracy of the rare symptoms and symptom combinations scales of the structured inventory of malingered symptomatology in three archival samples. *Law and Human Behavior, 44*(2), 167–177. https://doi.org/10.1037/lhb0000361.

Green, P. (2003). *Green's Word Memory Test for windows: User's manual*. Edmonton, Canada: Green's Publishing.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625–631. https://doi.org/10.1017/S1355617712000240.

Malcore, S., Schutte, C., Dyke, S., & Axelrod, B. (2015). The development of a reduced-item structured inventory of malingered symptomatology (SIMS). *Psychological Injury and Law, 8*(2), 95–99. https://doi.org/10.1007/s12207-015-9214-6.

Mazza, C., Orru, G., Burla, F., Monaro, M., Ferracuti, S., Colasanti, M., & Roma, P. (2019). Indicators to distinguish symptom accentuators from symptom producers in individuals with a diagnosed adjustment disorder: a pilot study on inconsistency subtypes using SIMS and MMPI-2-RF. *PLoS One, 14*, e0227113. https://doi.org/10.1371/journal.pone.0227113.

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.

Parks, A. C. Gfeller, J., Emmert, N., & Lammert, H. (2016). Detecting feigned postconcussional and posttraumatic stress symptoms with the structured inventory of malingered symptomatology (SIMS). *Applied Neuropsychology: Adult.* https://doi.org/10.1080/23279095.2016.1189426

Pearson. (2009). *Advanced clinical solutions for the WAIS-IV and WMS-IV: Clinical and interpretive manual*. San Antonio, TX: Pearson.

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: a follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*(3), 206–230. https://doi.org/10.1093/arclin/acw007.

Randolph, C. (1998). *Repeatable battery for the assessment of neuropsychological status manual*. San Antonio, TX: The Psychological Corporation.

Randolph, C. (2012). *Repeatable battery for the assessment of neuropsychological status update*. San Antonio, TX: The Psychological Corporation.

Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: the development of detection-based scales. *Behavioral Sciences & the Law, 32*(4), 455–466.

Silverberg, N. D., Wertheimer, J. C., & Fichtenberg, N. L. (2007). An effort index for the repeatable battery for the assessment of neuropsychological status (RBANS). *The Clinical Neuropsychologist, 21*(5), 841–854. https://doi.org/10.1080/13854040600850958.

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*(2), 183–189.

Spencer, R. J., Axelrod, B. N., Drag, L. L., Waldron-Perrine, B., Pangilinan, P. H., & Bieliauskas, L. A. (2013). WAIS-IV reliable digit span is no more accurate than age corrected scaled score as an indicator of invalid performance in a veteran sample undergoing evaluation for mTBI. *The Clinical Neuropsychologist, 27*(8), 1362–1372. https://doi.org/10.1080/13854046.2013.845243.

Spencer, R. J., Waldron-Perrine, B., Drag, L. L., Pangilinan, P. H., Axelrod, B. N., & Bieliauskas, L. A. (2017). Neuropsychological test validity in veterans presenting with subjective complaints of 'very severe' cognitive symptoms following mild traumatic brain injury. *Brain Injury, 31*(1), 32–38. https://doi.org/10.1080/02699052.2016.1218546.

Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. New York, NY: MHS Publishing.

van Beilen, M., Griffioen, B. T., Gross, A., & Leenders, K. L. (2009). Psychological assessment of malingering in psychogenic neurological disorders and non-psychogenic neurological disorders: relationship to psychopathology levels. *European Journal of Neurology, 16*(10), 1118–1123.

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.

Wechsler, D. (2008). Wechsler adult intelligence scale – Fourth edition (WAIS-IV). San Antonio, TX: The Psychological Corporation.

Widows, M. R., & Smith, G. P. (2005). *Structured inventory of malingered symptomatology: Professional manual*. Lutz, FL: Psychological Assessment Resources.