The Development of a Reduced-Item Structured Inventory of Malingered Symptomatology (SIMS)

Sylvia A. Malcore · Christian Schutte · Sarah A. Van Dyke · Bradley N. Axelrod

Abstract Effort assessment is a relevant area in neuropsychological assessment. There are well-established self-report measures, such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) FBS Symptom Validity Scale (FBS; Lees-Haley, English & Glenn, *Psychological Reports* 68: 203-210, 1991) and MMPI-2 Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant & Green, *Assessment* 14: 196, 2007). However, there is also a need for brief measures of symptom exaggeration. The Structured Inventory of Malingered Symptomatology (SIMS) appears to be a promising self-report measure of symptom validity (e.g., Smith & Burger, *Academy of Psychiatry and the Law* 25: 183-189, 1997). In order to understand the psychometric properties of the SIMS, a comprehensive item analysis was performed with the questionnaire. Items were removed due to invariable response or lack of relationship with the total SIMS. Receiver operating characteristics (ROC) curve analyses were completed using RBS and FBS Symptom Validity Scale (FBS) of the MMPI-2 with items grouping from the SIMS indices. The results found that limiting the number of items did not affect overall utility. The abbreviated indices on the SIMS had excellent discrimination with the neurologic impairment (NI) and affective disorder (AF) scales and good discrimination with psychosis (P) and amnesic disorder (AM) scales of the SIMS. The briefer measure allows for less time required to administer the measure, while still maintaining the integrity of the SIMS.

Keywords The Structured Inventory of Malingered Symptomatology (SIMS) · Symptom validity · MMPI-2

Sylvia A. Malcore and Sarah A. Van Dyke were both in training at John D. Dingell VA Medical Center at the time this work was performed.

S. A. Malcore · C. Schutte (✉) · S. A. Van Dyke · B. N. Axelrod
John D. Dingell VA Medical Center, Detroit, MI, USA
e-mail: dr.schutte@compneuropsych.com

Self-report measures of symptoms have been used as measures of effort indices to assess symptom validity, an important domain in neuropsychological testing (Heilbronner et al., 2009). Because evaluations are predicated on accurate presentation from the examinee, the validity of neuropsychological exams is dependent upon adequate effort (Wisdom, Callahan, & Shaw, 2010).

Mittenberg, Patton, Canyock, and Condit (2002) surveyed the American Board of Clinical Neuropsychology (ABCN) members that were in practice regarding the likelihood of symptom exaggeration. Estimated base rates of invalidity resulted from a number of diagnoses. The adjusted base rates for reported disorders included the following: mild head injury (41 %), moderate or severe head injury (9 %), fibromyalgia (39 %), neurotoxic disorders (29 %), electrical injury (26 %), Seizure disorder (9 %), and vascular dementia (2 %). Psychiatric conditions of pain or somatoform disorders (34 %), depressive disorders (16 %), anxiety disorders (14 %), and dissociative disorders (10 %) were also reported.

The base rates for exaggeration in the above disorders indicate the need to assess emotional overreport or exaggeration. Self-report measures, such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), can be used. The MMPI-2 is a well-validated measure with substantial literature, specifically there appears to be a support for the use of the FBS Symptom Validity Scale (sometimes also referred to as the SVS) as a symptom reporting validity scale (Lees-Haley, English, & Glenn, 1991; Larrabee, 1998; Greiffenstein, Backer, Gola, Donders, & Miller, 2002; Lewis, Simcox, & Berry, 2002; Larrabee, 2003; Greiffenstein, Back, Axelrod, Peck, & Gervais, 2004; Ross, Millis, Krukowski, Putnam, & Adams, 2004; Nelson, Hoelzle, Sweet, Arbisi, & Demakis, 2010). The FBS has been supported to assess exaggeration of symptoms in forensic settings, psychological injury litigants, and report of PTSD (Greiffenstein, et al., 2004). Nelson, Hoelzle, Sweet, Arbisi, and Demakis (2010)

96

Psychol. Inj. and Law (2015) 8:95–99

completed a meta-analysis of the FBS and pointed out that overall the literature is supportive of its use. Their study found the FBS to have a large composite effect size of .95 for comparison and overreporting groups.

A second scale, Response Bias Scale (RBS), has also been used to assess overreporting of additional somatic symptoms (Gervais, Ben-Porath, Wygant, & Green, 2007; Nelson, Sweet, & Heilbronner, 2007; Gervais, Ben-Porath, Wygant, & Green, 2008; Gervais, Ben-Porath, Wygant, & Sellbom, 2010; Wygant et al., 2010). Gervais, Ben-Porath, Wygant, and Green (2007) developed the RBS from the MMPI-2 using 28 items for a population involved in disability or forensic neuropsychological assessments. The researchers found support that the RBS had better incremental validity than the MMPI-2 scales of the FBS, Infrequency Scale (F), and Infrequency-Psychopathology (F_p) in predicting overreport on a number of validity tests. The authors suggest their study provides support for the use of the RBS as a measure in predicting incomplete effort on cognitive tests.

Whitney, Davis, Shepard, and Merman (2008) completed a validation study using retrospective data from 46 outpatient files from a VA Medical Center for patients evaluated for cognitive dysfunction. The study evaluated the relationship of the Test of Memory Malingering (TOMM) and MMPI-2 scales (RBS, FBS, Infrequency scale (F), Infrequency-Back scale (Fb) Infrequency Post-traumatic Stress Disorder Scale (Fptsd), Henry-Heilbronner Index (HHI)] Infrequency-Psychopathology Scale (Fp)). Their study indicated that only the RBS, Fb, and HHI were able to detect who passed versus who failed the TOMM. In addition to prediction of poor effort on cognitive functioning, the RBS has also been shown to detect exaggeration of depression or post-traumatic stress disorder.

Although symptom exaggeration measures are frequently used in neuropsychological assessment, they could also be useful in other general clinical settings given the base rates of exaggeration of other disorders. While there are measures that have been studied for their utility of detecting symptom exaggeration, the amount of time and/or training can be problematic. It would be beneficial for clinicians to have options of brief measures to detect possible exaggeration of symptoms (Edens, Poythress, & Watkins-Clay, 2007). There are limited options of brief stand-alone measures that assess for exaggeration and overreport with appropriate reliability and validity.

The SIMS appears to be a promising brief measure for identifying exaggeration of symptoms (e.g., Smith & Burger, 1997). The measure was developed with five indices of symptom overreport that the authors developed from a literature review of frequent feigned conditions, as well as adapted questions from other measures (e.g., MMPI) shown to distinguish individuals who are overreporting (Smith & Burger, 1997). The five indices include neurologic impairment (NI), affective disorders (AF), psychosis (P), low intelligence (LI), and amnesic disorders (AM). The authors described the development of the SIMS to include 200 original questions that were rated by clinical psychologists to determine which of the five categories of malingering the items matched too. The completed set of scales had interater reliability ranging ".76 (N) to .95 (AF), with a mean reliability of .84" (p. 186). The measure was developed by comparing seven simulation groups and one control group. The malingering group participants were provided a vignette in which they were asked to copy a condition and they were provided the instructions to only endorse the symptoms they believed were associated with the condition in order to avoid detection (Smith & Burger, 1997).

Smith and Burger (1997) found that the SIMS scales had an internal consistency with reliabilities for each index ranging from .80 to .88. The results of the study reflected that the total SIMS score was able to discriminate simulators and controls. The SIMS was compared to other self-report symptom validity measures (e.g., F and K scales of the MMPI-2), and each of the measures was able to discriminate between the controls and simulators.

Additional studies have continued to find support for the use of the SIMS. Edens, Otto, and Dwyer (1999) found that the SIMS was sensitive to symptom exaggeration, but it appeared to be at risk for a high number of false positives. The SIMS has been found to discriminate good from poor effort in forensic (e.g., Lewis et al., 2002) and veteran (e.g., Freeman, Powell, & Kimbrell, 2008) samples. Merckelbach and Smith (2003) found support for discriminate validity when comparing individual scores between the SIMS, BDI, and STAI-trait. People scoring within the upper 10 % of the later two measures typically were not above the cutoff of 16 on the SIMS. They also found that sensitivity was .93, and the specificity was .98 for the total scale of the SIMS. Finally, Wisdom, Callahan, and Shaw (2010) found utility in the SIMS for identifying potential symptom overreport in disability claims or personal injury lawsuits, but suggested using a higher cutoff in order to remove the potential for false positives and improve specificity.

Despite the preliminary support of the SIMS, the task contains 90 items and takes most examinees up to 30 minutes to complete it. The purpose of this study was to perform an item analysis of the SIMS in order to determine if a more streamlined measure could be extracted without significant loss of discriminability. The intent was to remove the items on the SIMS that do not contribute to identify individuals overreporting symptoms and compare the SIMS to well-established measures of symptom exaggeration.

Methods

Participants

Participants consisted of 158 consecutively evaluated individuals with mixed referral questions from an archival dataset from a large Midwestern Veterans Affairs medical center, of which 149 met the inclusion criteria. The sample had a mean age of 36.3 (SD=12.2), education was 13.0 (SD=1.9), and Full-Scale IQ from the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV FSIQ; The Psychological Corporation, 2008) in the average range (M=95.2 (SD=12.1). The sample was 92.1 % male. The sample was 69.0 % Caucasian, 28.6 % African American, and 2.4 % from other ethnicities.

Measures

The cognitive assessment included Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) Full-Scale IQ (FSIQ) for general intellectual ability, and the Test of Memory Malingering (TOMM) trial 2 was used to assess potential exaggeration of memory complaints through a forced choice paradigm (Tombaugh, 1996). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) was administered, from which FBS, RBS, Variable Response Inconsistency (VRIN), and True Response Inconsistency (TRIN) were extracted, to determine profile validity.

Patients with VRIN or TRIN T scores on the MMPI-2 that were higher than 85 were excluded from the sample (N=9) leaving a total useable sample of 149 participants. Individuals with FBS scores below 24 (N=96) were classified as within normal limits (WNL), and individuals with scores of 24 and above (N=53) were classified as overreporting (OR). For the first RBS cutoff, individuals with RBS scores below 12 (N=59) were classified as WNL, and individuals with scores of 12 and above (N=90) were classified as OR. For the second RBS cutoff, individuals with RBS scores below 17 (N=114) were classified as WNL, and individuals with scores of 17 and above (N=35) were classified as OR.

Results

Item Extraction

Correlation coefficients were computed among the five indices of the SIMS and trial 2 of the TOMM, which were consistently low, ranging from −.26 to −.33 with the five indices. Additionally, correlation coefficients between the five SIMS indices and WAIS-IV FSIQ were also consistently low, ranging from −.11 to −.25, with the exception of the low

intelligence (LI) index, which was correlated at −.34. Although the correlations between indices and the TOMM and/or Full-Scale IQ were statistically significant (p<.05), the strength of the relationship was weak (see Table 1). As suggested by Millis and Volinsky (2001), statistically significant relationships only indicate it is different from 0 and not necessarily useful diagnostically. Therefore, the SIMS does not appear to be measuring poor effort on cognitive functioning or intellectual ability.

An item analysis was completed to assess contribution of individual items to the total score. We employed two rules to determine which of the 75 individual items to remove. First, individual items with a corrected item-total correlation of less than r=.3 were removed due to a lack of strength in the relationship with the measure. A total of 37 items was removed from NI (n=3), AF (n=9), P (n=10), LI (n=13), and AM (n=2). From the remaining 39 items, items in which 95 % or greater of the respondents provided the same response were removed as they were largely invariable. One additional item was removed from the P scale. Indices were then left with the following number of items: NI (n=12), AF (n=6), P (n=4), AM (n=13), and LI (n=2). Since only two items remained in the LI scale, the scale was removed, resulting in a total of 36 items (see Appendix).

The indices of the SIMS were then recorrelated with the TOMM and Full-Scale IQ. Correlations continued to be low, ranging from −.30 to .26 for the TOMM and −.31 to −.10. Therefore, we elected to evaluate the SIMS against the MMPI-2 FBS and RBS, as they are also self-report measures of exaggeration (see Table 1).

Comparisons of New SIMS to Standard SIMS

Correlations between the original SIMS indices with both FBS and RBS demonstrated that NI, AF, P, and AM were significantly related to FBS and RBS at p<.01. The LI scale was significantly related to RBS, but not to FBS (see Table 2). Correlations between the abbreviated SIMS indices and FBS and RBS demonstrated that NI, AF, P, and AM were also significantly related to FBS and RBS at p<.01. Receiver

Table 1 Correlation coefficients for relations among measures of TOMM, FSIQ, and SIMS indices

Measure	SIMS NI	SIMS AF	SIMS P	SIMS LI	SIMS AM
1. TOMM					
Original	−.30**	−.26**	−.29**	−.33**	−.26**
Abbreviated	−.30**	−.25**	−.27**	—	−.26**
2. FSIQ					
Original	−.25**	−.24**	−.20*	−.34**	−.11
Abbreviated	−.29**	−.21*	−.31**	–	−.10

*p<.05; **p<.01

98

Psychol. Inj. and Law (2015) 8:95–99

Table 2 Correlation coefficients for relations among measures of MMPI-2 and SIMS indices

Measure	SIMS NI	SIMS AF	SIMS P	SIMS LI	SIMS AM
1. FBS					
Original	.42**	.28**	.33**	.10	.37**
Abbreviated	.42**	.27**	.29**	–	.35**
2. RBS					
Original	.64**	.52**	.39**	.21**	.63**
Abbreviated	.66**	.50**	.55**	–	.62**

*p<.05; **p<.01

operating characteristics (ROC) curve analyses were conducted to assess how well each of the original and the abbreviated SIMS scales and total scores discriminated between individuals who were classified as overreporting (OR) on the FBS and the two different RBS cutoff scores (see Table 3). On the prediction of group membership with the FBS, the original SIMS indices had a total achieved area under the curve (AUC) of .74 for NI, .62 for AF, .62 for P, .58 for LI, .73 for AM, and .72 for the total score. On the prediction of group membership with the FBS, the abbreviated SIMS indices and total had an achieved AUC of .73 for NI, .61 for AF, .65 for P, .73 for AM, and .75 for the total score. As the ROC curve analyses did not achieve an AUC of greater than .75, the RBS was used to determine optimal cut scores. On the prediction of group membership with the RBS (>12 cut score), the original

Table 3 ROC for original and abbreviated SIMS indices

Scale	FBS	RBS12	RBS16
1. SIMS NI			
Original	.74	.79	.82
Abbreviated	.73	.80	.84
2. SIMS AF			
Original	.62	.71	.77
Abbreviated	61	.73	.77
3. SIMS P			
Original	.62	.62	.69
Abbreviated	.65	.69	.79
4. SIMS LI			
Original	.58	.55	.58
Abbreviated	–	–	–
5. SIMS AM			
Original	.73	.83	.81
Abbreviated	.73	.83	.81
6. SIMS total			
Original	.72	.84	.85
Abbreviated	.75	.87	.88

RBS12=cut score of >12 and RBS16=cut score of >16

SIMS indices had a total achieved area under the curve (AUC) of .79 for NI, .71 for AF, .62 for P, .55 for LI, .83 for AM, and .84 for the total score. On the prediction of group membership with the RBS (>12 cut score), the abbreviated SIMS indices had a total achieved area under the curve (AUC) of .80 for NI, .73 for AF, .69 for P, .83 for AM, and .87 for total. On the prediction of group membership with the RBS (>16 cut score), the original SIMS indices had a total achieved area under the curve (AUC) of .82 for NI, .77 for AF, .69 for P, .58 for LI, .81 for AM, and .85 for the total scores. On the prediction of group membership with the RBS (>16 cut score), the abbreviated SIMS indices had a total achieved area under the curve (AUC) of .84 for NI, .77 for AF, .79 for P, .81 for AM, and .88 for the total score.

In determining optimal cut scores for the abbreviated indices and total score of the SIMS, RBS (both >12 and >16 cutoff) was used as the AUC between indices, and FBS was not greater than .75. Cut scores were determined by selecting the score that produced a specificity at or near .90 to reduce the occurrence of false positive scores in using this brief self-report measure of reporting style. The cut scores based on the RBS (>12) were NI >5 (sensitivity=.46, specificity=.93), AF>4 (sensitivity=.37, specificity=.86), P>2 (sensitivity=..18, specificity=.98), AM>6 (sensitivity=..61, specificity=.92), and total score >12 (sensitivity=.64, specificity=.92). The cut scores based on the RBS (>16) were NI>6 (sensitivity=.49, specificity=.93), AF>5 (sensitivity=.23, specificity=.91), P>2 (sensitivity=.37, specificity=.97), AM>9 (sensitivity=.34, specificity=.92), and total score >17 (sensitivity=.60, specificity=.90).

Discussion

The assessment of symptom exaggeration is important in a variety of clinical settings. Although the original SIMS appears to offer a promising alternative, current study suggests that a briefer version resulted in no loss of efficacy. Both the original and the abbreviated SIMS performed similarly to well-validated MMPI-2 indices, with the abbreviated SIMS performing slightly better than the original SIMS for some of the index scales. These results suggest that the SIMS can be shortened while maintaining the integrity of the measure since ROC curve analyses reflect similar AUC between the original and abbreviated SIMS. The analyses in the current study reduced the SIMS to a 36-item scale with four of the original indices (NI, AF, P, and AM).

The total score index of the abbreviated SIMS was not stronger than the abbreviated NI or abbreviated AM, suggesting that the addition of the items does not add to the ability of the measure to differentiate performance on RBS. The results suggest that use of one of the stronger indices, using only 14 items, is as effective as the complete 90-item questionnaire.

The findings of the study are encouraging the use of abbreviated item SIMS. However, due to the exploratory nature of the study, the results will need to be replicated before it is applied to clinical settings.

Appendix

Items removed due to an corrected item-total correlation of less than $r=.3$ or if 95 % or greater of the respondents provided the same response.

NI	44, 54, 74
AF	2, 16, 23, 24, 32, 43, 52, 55, 72
P	3, 8, 10, 28, 34, 38, 42, 48, 57, 69, 65
LI	4, 7, 11, 14, 21, 46, 56, 58, 63, 67, 68, 73, 75
AM	9, 40

References

Butcher, J. N., Dalstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the administration and scoring of the MMPI-2*. Minneapolis: University of Minnesota Press.

Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the structured inventory of malingered symptomatology in identifying persons motivated to malinger psychopathology. *The Journal of the American Academy of Psychiatry and the Law, 27*(3), 387–396.

Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: a comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment, 88*(1), 33–42.

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research, 158,* 374–380.

Gervais, R., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14*(2), 196–208.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *The Clinical Neuropsychologist, 22,* 1061–1079.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2-RF over-reporting scales and RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology, 25,* 274–284.

Greiffenstein, M. F., Baker, W. J., Axelrod, B., Peck, E. A., & Gervais, R. (2004). The Fake Bad Scale and MMPI-2 F-family in detection of implausible psychological trauma claims. *The Clinical Neuropsychologist, 18,* 573–590.

Greiffenstein, M. F., Baker, W. J., Gola, T., Donders, J., & Miller, L. (2002). The Fake Bad Scale in atypical and severe closed head injury litigants. *Journal of Clinical Psychology, 58*(12), 1591–1600.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & conference participants. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23,* 1093–1129.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist, 12*(2), 179–188.

Larrabee, G. J. (2003). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist, 17*(1), 54–68.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychological Reports, 68,* 203–210.

Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment, 14*(2), 170–176.

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,* 145–152.

Millis, S. R., & Volinsky, C. T. (2001). Assessment of response bias in mild head injury: beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology, 23*(6), 809–828.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094–1102.

Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A., & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 symptom validity scale (FBS): verified utility in forensic practice. *The Clinical Neuropsychologist, 24,* 701–724.

Nelson, N. W., Sweet, J. J., & Heilbronner, R. L. (2007). Examination of the new MMPI-2 Response Bias Scale (Gervais): relationship with MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 29*(1), 67–72.

Ross, S. R., Millis, S. R., Krukowski, R. A., Putnam, S. H., & Adams, K. M. (2004). Detecting incomplete effort on the MMPI-2: an examination of the Fake-Bad Scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology, 26*(1), 115–124.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *The Journal of the American Academy of Psychiatry and the Law, 25*(2), 183–189.

The Psychological Corporation. *Wechsler Adult Intelligence Scale, fourth edition manual*. Author, San Antonio, TX, 2008.

Tombaugh, T. N. (1996). *Test of memory malingering*. Toronto, Canada: MultiHealth Systems.

Whitney, K. A., Davis, J. J., Shepard, P. H., & Herman, S. M. (2008). Utility of the response bias scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology, 23,* 777–786.

Wisdom, N. M., Callahan, J. L., & Shaw, T. G. (2010). Diagnostic utility of the structured inventory of malingered symptomatology to detect malingering in a forensic sample. *Achieves of Clinical Neuropsychology, 25,* 118–125.

Wygant, D. B., Sellbom, M., Gervais, R. O., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2010). Further validation of the MMPI-2 and MMPI-2-RF Response Bias Scale: findings from disability and criminal forensic settings. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 22*(4), 745–756.