



Pergamon

Archives of Clinical Neuropsychology
18 (2003) 145–152

Archives
of
CLINICAL
NEUROPSYCHOLOGY

Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering

Harald Merckelbach^{a,*}, Glenn P. Smith^b

^a*Department of Experimental Psychology, Maastricht University, P.O. Box 616,
6200 MD, Maastricht, The Netherlands*

^b*James A. Haley Veterans' Hospital, Tampa, FL, USA*

Accepted 20 October 2001

Abstract

This article addresses the psychometric properties of the Dutch translation of the Structured Inventory of Malingered Symptomatology (SIMS) when administered to undergraduate psychology students as well as psychiatric inpatients. Findings show that this SIMS version possesses good test–retest reliability and internal consistency. Also, simulation findings indicate that undergraduate students instructed to simulate pathology display higher SIMS scores than either normal controls or psychiatric inpatients. Data pooled over several samples ($n = 298$) yielded sensitivity, specificity, and positive predictive power (PPP) rates that were all relatively high (≥ 0.90). All in all, our findings provide a basis for cautious optimism regarding the usefulness of the SIMS as a screening tool for malingering.

© 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd. All rights reserved.

Keywords: SIMS; Malingering; Depression; Fantasy proneness; Diagnostic accuracy

1. Introduction

The Structured Inventory of Malingered Symptomatology (SIMS; [Smith, 1997](#); [Smith & Burger, 1997](#)) is a self-report measure designed to screen for malingering of psychiatric symptoms (e.g., depression and psychosis) and/or cognitive impairments (e.g., low intelligence and memory complaints). The SIMS consists of 75 dichotomous (i.e., true–false) items that

* Corresponding author. Tel.: +31-43-3881908; fax: +31-43-3884196.

E-mail address: h.merckelbach@psychology.unimaas.nl (H. Merckelbach).

can be grouped into five subscales, each subscale containing 15 items. Subscales tap malingered symptoms in several areas. More specifically, they focus on the following domains: low intelligence (LI), affective disorders (AF), neurological impairment (N), psychosis (P), and amnesic disorders (AM). Strategies used to detect deviant or malingered response patterns include endorsement of bizarre experiences (e.g., “Sometimes my muscles go limp for no apparent reason so that my arms and legs feel as though they weigh a ton” from the N scale), highly atypical symptoms (e.g., “At times, I am so depressed I welcome going to bed early to ‘sleep it off’” from the AF scale), and Ganser-like (i.e., approximate) answers (e.g., “If you have US\$1.50 and I take fifty cents away, you will have 75 cents left” from the LI scale).

A number of analog studies have looked at the accuracy with which the SIMS detects malingered symptomatology (Edens, Otto, & Dwyer, 1999; Rogers, Hinds, & Sewell, 1996; Smith & Burger, 1997). These studies found relatively high sensitivity and specificity rates, which suggest that the SIMS is a promising instrument. For example, Smith and Burger (1997) instructed undergraduates to feign in a convincing manner LI, AF, N, P, or AM. Performance of the malingering groups on the total SIMS and the SIMS subscales were compared to those of a control group. Using a cutoff of 14 for the SIMS total score, the authors found that 96% of the malingerers (sensitivity) and 88% of the controls (specificity) were correctly classified. However, sensitivity and specificity rates of the separate SIMS subscales were generally lower, which lead the authors to conclude that the SIMS total score is a good overall indicator of malingering, while SIMS subscale scores only provide qualitative information about the type of symptoms that individuals try to feign.

Edens et al. (1999) reached a similar conclusion. In their study, students completed the SIMS twice. On one occasion, students were instructed to answer SIMS items honestly, while on the other occasion they were told to convincingly fake psychotic, depressive, or cognitive dysfunction symptoms. Again, high sensitivity and specificity rates were found for SIMS total scores (96 and 91%, respectively), whereas those of the subscale scores remained relatively low. Edens et al. also noted that specificity of the SIMS total scale dropped for those individuals who reported high levels of current psychopathology (as measured by the SCL-90). That is, using a cutoff of 14, the SIMS total scale misclassified 22% of the high SCL-90 nonmalingerers. Accordingly, the authors conclude that “higher cutoff scores ultimately may be needed to increase specificity” (Edens et al., 1999, p. 395).

In their study on adolescent offenders, Rogers et al. (1996) had their subjects fill in the SIMS twice: once under honest and once under feigning conditions. The authors calculated positive predictive power (PPP) and negative predictive power (NPP) for the SIMS. PPP refers to the probability that an individual with a score that exceeds the cutoff is a malingerer, while NPP refers to the probability that an individual with a score below the cutoff is an honest responder. Using a cutoff point of 16 for the SIMS total scale, Rogers et al. found a PPP of 0.87 and a NPP of 0.62. The authors argued that the relatively high PPP rate is encouraging and shows that the SIMS might, indeed, be an effective screening device. On the other hand, the relatively low NPP rate indicates that at least in this sample, a high frequency of false-negatives (malingerers who were classified as honest responders) occurred. This finding argues for the use of the SIMS as a trigger for a full evaluation—not for the actual determination of malingering. The sensitivity and specificity rates for the SIMS that can be derived from the Rogers et al.

data (0.43 and 0.94, respectively) underline this point. Note that compared to the results of other studies (e.g., Edens et al., 1999; Smith & Burger, 1997), the sensitivity is unexpectedly low.

So far, data that have been gathered about the diagnostic accuracy of the SIMS support its use as a screening tool. However, two critical notes can be raised. First, little is known about SIMS scores of individuals with severe psychopathology or those who are at risk for psychopathology. The findings reported by Edens et al. (1999) suggest that such individuals may have raised SIMS scores, but the question is, of course, to what extent their scores exceed the recommended cutoff of 16. Secondly, base rates of malingerers within research samples are probably considerably higher than those found in clinical populations. Some authors have emphasized that sensitivity, specificity, PPP, and NPP rates for screening instruments like the SIMS are not very informative unless these base rates are taken into account (see, for a thorough analysis of this point, Rosenfeld, Sands, & Van Gorp, 2000). In the studies cited above, the base rate of malingering always circled around 50%. On the other hand, a survey among forensic experts (Rogers, Sewell, & Goldstein, 1994) estimated the base rate of malingering in forensic settings to be only 17%.

With these two points in mind, we conducted the current study. It relied on a Dutch translation of the SIMS that was administered to several samples. Apart from the standard psychometric information (i.e., test–retest stability and Cronbach α coefficients), we obtained SIMS scores of individuals scoring high on depression and/or trait anxiety. We also compared SIMS scores of instructed malingerers to those of psychiatric patients and normal controls. Finally, sensitivity, specificity, PPP, and NPP rates were calculated for the pooled data in which malingering was a phenomenon with a low base rate frequency.

2. Method

2.1. Participants

The current study involved three samples (see Table 1). Sample 1 consisted of 24 undergraduate psychology students (18 women). Their mean age was 21.3 years (S.D. = 1.8 years).

Table 1
Size, mean age, and mean SIMS total scores of Samples 1–3

Sample	<i>n</i> (women)	Age (S.D.)	SIMS total score (S.D.)
1	24 (18)	21.3 (1.8)	1:3.8 (2.3), 2:3.4 (3.4)
2	182 (149)	19.3 (1.9)	5.8 (3.9)
3			
Controls	25 (17)	20.1 (2.1)	5.1 (4.5)
Amnesia	28 (19)	19.8 (2.0)	34.8 (9.9)
Schizophrenia	14 (13)	19.9 (1.5)	37.7 (14.3)
Neurological disease	15 (12)	20.1 (1.2)	30.3 (11.0)
Patients	10 (3)	38.1 (4.6)	11.4 (5.8)

Sample 2 consisted of 182 (149 women) undergraduate psychology students with a mean age of 19.3 years (S.D. = 1.9 years). Sample 3 consisted of 82 (60 women) undergraduate psychology students with a mean age of 20.1 years (S.D. = 2.0 years) and 10 psychiatric inpatients (three women) with a mean age of 38.1 years (S.D. = 4.6 years). All subjects volunteered to participate in the current study. Psychiatric patients were randomly recruited inpatients from a psychiatric hospital. They volunteered to complete the SIMS after they had given informed consent. Their diagnoses varied from substance abuse and borderline personality disorder to bipolar disorder and schizophrenia.

2.2. Materials and procedure

All participants completed a Dutch version of the SIMS (Smith & Burger, 1997). The original SIMS items were translated and back translated by a native speaker in order to remove ambiguities in the Dutch version. Also, typical American details (e.g., US\$) were replaced by Dutch equivalents. SIMS total scores range from 0 to 75. The cutoff score of 16 recommended by Rogers et al. (1996) was used.

Participants of Sample 1 completed the Dutch version of the SIMS on two occasions, with a 3-week interval in between. During both occasions, participants were instructed to respond honestly to the SIMS items. In this way, test–retest stability could be calculated. Participants were tested in small groups of six persons.

Participants in Sample 2 were also asked to answer the SIMS items honestly. In addition, they completed the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the trait anxiety scale of the Spielberger State-Trait Anxiety Inventory (STAI-trait; Spielberger, Gorsuch, & Lushene, 1970), and the Creative Experiences Questionnaire (CEQ; Merckelbach, Horselenberg, & Schmidt, 2001).

The BDI is a 21-item self-report scale measuring behavioral manifestations of depression. Items are rated on 4-point scales (range: 0–3) and scores are then summed to obtain a total BDI score (range: 0–63), with higher scores indicating higher levels of depression. The STAI-trait consists of 20 items that address habitual positive and negative feelings related to anxiety. Items are scored on 4-point scales (range: 1–4) and after recoding of the positive items, a total score can be obtained by summing across items (range: 20–80). Higher scores indicate higher trait anxiety. The CEQ is a 25-item self-report scale of fantasy proneness. The false–true items refer to deep involvement in daydreaming, make-belief, and imagination. Scores are summed to obtain a total CEQ score (range: 0–25), with higher scores reflecting higher levels of fantasy proneness. Participants of Sample 2 completed all scales during a mass testing session.

Sample 3 only completed the SIMS. However, there were five different conditions. They were undergraduates ($n = 25$) responding honestly, patients ($n = 10$) responding honestly, undergraduates ($n = 28$) simulating amnesia, undergraduates ($n = 14$) simulating schizophrenia, and undergraduates ($n = 15$) simulating neurological problems. In each of the simulation conditions, participants were presented with a detailed vignette outlining a particular set of complaints made in a forensically relevant context (e.g., litigation) and instructing the person to present himself or herself in this fashion and to do so in manner that would convince clinicians. Participants of Sample 3 were tested individually.

3. Results

3.1. Reliability of the SIMS

Table 1 summarizes SIMS total scores of the three samples. As can be seen, scores on the first and the second test occasions of the honestly responding participants in Sample 1 remained practically the same [$t(23) = 1.1, P > .30$]. The test–retest correlation was $.72 (P < .01)$, indicating that SIMS scores are relatively stable across time. The Cronbach α coefficient for the SIMS total scale obtained in Sample 2 was $.72$, which supports the internal consistency of the SIMS total scale. Cronbach α coefficients for the subscales were considerably lower and varied between $.24$ (LI) and $.59$ (AF).

3.2. Construct validity

Mean BDI, STAI-trait, and CEQ scores of Sample 2 were 5.8 (S.D. = 5.4), 39.3 (S.D. = 9.2), and 7.6 (S.D. = 3.9), respectively. Mean SIMS total score for the honestly responding participants of Sample 2 was 5.8 (S.D. = 3.9). SIMS total scores were significantly correlated with depression as indexed by the BDI ($r = .64, P < .01$), trait anxiety as measured by the STAI-trait ($r = .55, P < .01$), and fantasy proneness as measured by the CEQ ($r = .33, P < .01$). The link between SIMS and CEQ supports the convergent validity of the SIMS inasmuch as the SIMS, like the CEQ, is sensitive to the tendency to endorse bizarre and atypical items. The connections between SIMS, depression, and trait anxiety argue against the discriminant validity of the SIMS, because these links suggest that the SIMS is sensitive to real psychopathology. On the other hand, 3 out of the 182 participants (1.7%) in Sample 2 had a SIMS total score exceeding the cutoff of 16. Table 2 shows numbers of participants in the upper 10% of the BDI, STAI-trait, and CEQ distributions with SIMS total scores above this cutoff point. As can be seen, these rates were relatively low. This implies that notwithstanding the significant correlations between SIMS, BDI, and STAI-trait, false-positive rates for the SIMS remain relatively low.

SIMS total scores of controls, instructed simulators, and patients in Sample 3 are given in Table 1. As another exploration of the construct validity of the SIMS, a one-way analysis of variance (ANOVA) was performed on these data. Results indicated that the three groups differed with regard to their SIMS total scores [$F(2, 88) = 87.7, P < .01$]. Follow-up t tests showed that instructed simulators had higher SIMS total scores than patients [$t(65) = 6.2, P < .01$], who in turn had higher scores than normal controls [$t(32) = 3.2, P < .01$].

To examine whether SIMS subscales were sensitive to the particular symptoms that were feigned, a 3×5 ANOVA with repeated measures on the last factor was conducted. The three

Table 2

Number (%) of participants of full Sample 2 ($n = 182$) with SIMS total score >16 and number (%) of participants scoring in the upper 10% of the BDI (depression), STAI-trait (trait anxiety), and CEQ (fantasy proneness) distribution with SIMS total scores >16

Total sample ($n = 182$)	BDI >13 ($n = 19$)	STAI-trait >52 ($n = 19$)	CEQ >12 ($n = 19$)
3 (1.7%)	3 (1.7%)	2 (1.1%)	1 (0.6%)

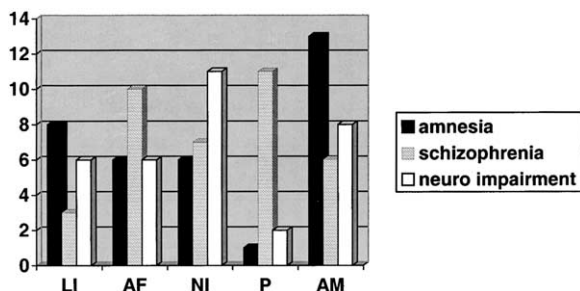


Fig. 1. SIMS subscale scores of undergraduates who were instructed to simulate amnesia ($n = 28$), schizophrenia ($n = 14$), or neurological impairments ($n = 15$).

factors were the simulation conditions: amnesia, schizophrenia, and neurological symptoms. The five factors were the SIMS subscales: LI, AF, N, P, AM. Although there were no overall differences between the three simulation conditions [$F(2, 54) = 1.6, P = .22$], the simulation Condition \times Subscale interaction was highly significant [$F(8, 246) = 33.4, P < 0.01$]. As can be seen in Figure 1, the feigned amnesia group had the highest score on the SIMS amnesia subscale, the feigned schizophrenia group had the highest score on the SIMS psychosis subscale, whereas the feigned neurological impairment group had the highest score on SIMS neurological impairment subscale. This pattern supports the idea (Smith & Burger, 1997) that SIMS subscales provide qualitative information as to the particular syndrome that people try to simulate.

3.3. Predictive accuracy

SIMS data of Samples 1–3 were pooled and the following four indices of predictive accuracy were calculated (see Rosenfeld et al., 2000): sensitivity, specificity, PPP, and NPP. The pooled sample consisted of 298 participants of whom 57 (19%) were known to feign symptoms on the SIMS. Note that the base rate of malingering comes close to the 17% reported by Rogers et al. (1994). Using a cutoff point of 16 for the SIMS total score, accuracy indices can be calculated from the data presented in Table 3. As can be seen, 93% of the malingerers were correctly identified by the SIMS (sensitivity), while 98% of the honest responders were correctly

Table 3
Classification and accuracy indices for the pooled data ($n = 298$)

	Actual: malingering	Actual: honest
Predicted: malingering	53 (TP)	6 (FP)
Predicted: honest	4 (FN)	235 (TN)
Sensitivity	.93	
Specificity	.98	
PPP	.90	
NPP	.98	

Sensitivity = $TP/(TP+FN)$; specificity = $TN/(FP+TN)$; PPP = $TP/(TP+FP)$; NPP = $TN/(FN+TN)$.

classified (specificity). The probability that a person with a SIMS total score exceeding the cutoff of 16 is a malingerer (PPP) was found to be 90%, while the likelihood that a person with a SIMS score below the cutoff is a nonmaligner (NPP) was found to be 98%. All in all, these accuracy rates are encouraging.

4. Discussion

The main findings of the current study can be catalogued as follows. To begin with, stability and internal consistency of the SIMS were found to be satisfactory, which supports the reliability of this instrument. Secondly, robust correlations between SIMS, depression, and trait anxiety were found. This is in line with [Edens et al. \(1999\)](#) who noted a certain amount of overlap between SIMS and scales that are thought to tap real psychopathological symptoms (e.g., SCL-90). On the basis of this overlap, Edens et al. (p. 395) concluded that “genuinely symptomatic persons are at risk for being identified as malingering on the SIMS.” However, the current results suggest that at least in undergraduate samples, the overlap between SIMS, depression, and trait anxiety pertains to the lower portions of the SIMS distribution. That is, SIMS scores of persons scoring in the upper 10% of the BDI or STAI-trait were rarely found to exceed the cutoff point of 16. In passing, it should be noted that mean BDI and STAI-trait scores of these persons were well within the clinical range (e.g., [Clark, 1992](#); [Cloitre & Liebowitz, 1991](#)). One could counter that in the high depression subsample, 3 out of 19 participants had a SIMS score exceeding the cutoff (see [Table 2](#)). Following this line of reasoning, the false-positive rate for this particular subsample would be 16%, which is considerably worse than the false-positive rate of 2% indicated by the specificity of .98 for the entire sample. Although this is a valid point, it should be qualified in two ways. To begin with, as malingering might involve not only fabrication but also exaggeration of symptoms, it is possible that in the subsample of high BDI participants, some individuals exaggerated their depressive symptoms. In that case, the true false-positive rate for this subsample would be lower than 16%. Furthermore, it should be stressed that the SIMS intends to be a screening device. Screening devices require a more thorough follow-up evaluation and therefore it is acceptable when such devices are relatively liberal when they “rule in” potential malingerers. Nevertheless, the overlap between SIMS and self-reported psychopathology such as depression warrants further study.

Thirdly, the current study found evidence to support the validity of the SIMS. More specifically, a modest but significant correlation emerged between fantasy proneness and SIMS. This correlation was anticipated since fantasy proneness is accompanied by a positive response bias when answering odd items ([Merckelbach, Muris, Horselenberg, & Stougie, 2000](#)). Furthermore, instructed malingerers had significantly higher SIMS total scores than either honestly responding controls or psychiatric inpatients. Although Cronbach α coefficients for separate SIMS subscales were low, the subscales were found to be sensitive to the type of symptoms that instructed malingerers tried to feign. This supports the qualitative use of subscales, as recommended by [Smith and Burger \(1997\)](#).

Fourthly, the accuracy indices derived from the pooled SIMS data were impressive. Using a cutoff of 16, sensitivity and specificity proportions for the SIMS total scale were .93 and .98, respectively. These rates correspond with those reported by [Edens et al. \(1999\)](#) and [Smith](#)

and Burger (1997). PPP and NPP were found to be .90 and .98. Interestingly, these rates are considerably above the PPP and NPP rates reported by Rogers et al. (1996), although the base rate of malingering in our pooled sample was unfavorable compared to that used in the Rogers et al. study (i.e., 19% vs. 50%, respectively). As Rosenfeld et al. (2000) point out, the interpretation of accuracy indices heavily depends on base rate frequencies. Even if one would assume that the base rate of malingering is 5%, our selectivity and sensitivity rates would imply rather encouraging PPP and NPP probabilities (.71 and .99, respectively). On the other hand, two potential limitations of the current study deserve some comment. Firstly, in our simulation conditions, malingerers did not receive positive or negative incentives. Meanwhile, there are good reasons to believe that such incentives might affect SIMS scores (Rogers & Cruise, 1998). Secondly, our patient sample was small and diagnostically heterogeneous. Given the correlations between depression, trait anxiety, and SIMS, large-scale clinical studies are required to shed more light on the diagnostic accuracy of the SIMS in genuinely symptomatic individuals.

To sum up, then, the current findings found further evidence for the reliability and validity of the SIMS. Especially, the high PPP and NPP rates provide a basis for cautious optimism regarding the usefulness of the SIMS as a screening tool for malingering.

References

- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571.
- Clark, D. A. (1992). Depressive anxious and intrusive thoughts in psychiatric inpatients and outpatients. *Behaviour Research and Therapy*, 30, 93–102.
- Cloitre, M., & Liebowitz, M. R. (1991). Memory bias in panic disorder: An investigation of the cognitive avoidance hypothesis. *Cognitive Therapy and Research*, 15, 371–386.
- Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malingering psychopathology. *Journal of the American Academy of Psychiatry and the Law*, 27, 387–396.
- Merckelbach, H., Horselenberg, R., & Schmidt, H. (2001). The creative experiences questionnaire: A brief self-report measure of fantasy proneness. *Personality and Individual Differences*, 31, 987–995.
- Merckelbach, H., Muris, P., Horselenberg, R., & Stougie, S. (2000). Dissociation, reality monitoring, and response bias. *Personality and Individual Differences*, 28, 49–58.
- Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, 22, 273–285.
- Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment*, 67, 244–257.
- Rogers, R., Sewell, K. W., & Goldstein, A. M. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior*, 18, 543–552.
- Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, 15, 349–359.
- Smith, G. P. (1997). Assessment of malingering with self-report instruments. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 351–370). New York: Guilford.
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the Academy of Psychiatry and the Law*, 25, 180–183.
- Spielberger, C. D., Gorsuch, R. K., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.