Detection of Malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS)

Glenn P. Smith, PhD, and Gary K. Burger, PhD

This article discusses the development and validation of a paper and pencil screening measure, the Structured Inventory of Malingered Symptomatology (SIMS), designed to detect malingering. Test items were constructed from a combination of revised validity questions from existing instruments and characteristics of malingerers noted by existing research. Items were organized on one of five subscales by experienced clinical psychologists. College students (N = 476) were assigned to one of seven simulation conditions (i.e., psychosis, amnesia, neurologic impairment, mania, depression, low intelligence, and "fake bad") or an honestly responding group. All subjects were administered the SIMS, the F and K scales of the MMPI, 16PF Faking Bad scale, and portions of the malingering scale. The SIMS total score demonstrated the highest sensitivity rating (95.6%) for detection when compared with the other validity indices. Suggestions concerning further research using the SIMS as well as its potential utility in a complete evaluation process are discussed.

Individuals wishing to present themselves as suffering from psychological or physical condition can alter the nature of their self-reported symptoms in at least two general ways: by intentionally fabricating nonexistent symptoms or by exaggerating genuine ones. When the motivation for this type of impression management occurs in the light of secondary gain, such

invalid responding is referred to as malingering. In general, psychometric approaches to detect this condition have relied on the use of either validity scales that are incorporated into an assessment device or through a malingering-specific instrument.

The Minnesota Multiphasic Personality Inventory (MMPI)[1] is an example of an assessment device with validity indices. One of these indices is the F scale, that uses endorsement of low frequency items to demonstrate an atypical response style.[2] The use of the F scale for the detection of malingering, one of the styles

Dr. Burger is affiliated with the University of Missouri-St. Louis; Dr. Smith with the James A. Haley Veterans Affairs Hospital, Tampa, FL. Address correspondence to: Glenn P. Smith, PhD, James A. Haley Veterans Affairs Hospital, 13000 Bruce B. Downs Blvd., Tampa, FL 33162.

of atypical responding, has produced mixed results.[2] Attempts to derive more sensitive cutoff values have been limited by the considerable overlap between the scores of malingerers and genuine diagnostic groups.[3] Roman et al.,[4] for example, found that using the F scale correctly classified all malingerers within a forensic population but at the cost of misidentifying 41 percent of the individuals suffering from psychosis. Further complicating the use of the F scale, researchers have argued for different cutoff values within distinct populations.[5-7] One inherent difficulty with using the F scale to detect malingered profiles is the fact that the items are not sensitive to certain types of feigned conditions (e.g., malingered memory problems).

Another objective personality measure, Cattell's 16 Personality Factor Questionnaire (16PF) Form A,[8] also includes a validity index, the Faking Bad (FB) scale. The test originators proposed a cutoff of 6 or higher on this 15-item scale that reportedly detected 94% of individuals attempting to "fake bad." Krug[9] considered the reliability and distribution of this validity scale across a large sample and concluded that this cutoff "seems about right" (p. 156). However, a review of the literature revealed no controlled studies to confirm this criterion. As a result, the efficacy of the FB scale for detecting malingered profiles is unknown.

The M Test[10] was among the first attempts to develop a device specifically to detect the presence of malingering. Investigations into the usefulness of the instrument have yielded poor hit rates when using the authors' original criteria.[11, 12]

Hankins et al.,[13] in fact, found that response patterns appeared to reflect more the severity of cognitive impairment of the subject than the presence of malingering. In an attempt to improve its discriminability, Rogers et al.[14] developed a revised scoring system for the M Test (i.e., rule-out and rule-in scales). Their subsequent investigation resulted in the successful identification of more than 80 percent of the potential malingerers within correctional and inpatient psychiatric samples. However, Smith et al.[15] failed to cross-validate these results in a similar population.

The malingering scale (MS)[16] is another malingering-specific instrument. The MS is composed of two scales, the malingering retardation (MgR) scale and the malingering insanity (MgI) scale. In a cross-validation study,[7] the MS, in combination with the MMPI and Bender-Gestalt, correctly classified substance abusers instructed to fake insanity from psychiatric inpatients 97.6 percent of the time. However, generalizability was limited by differential incentives between malingering groups, the analogue nature of the design, and reduced variability in IQ scores within the experimental sample. Further, the test has certain logistical drawbacks such as its length (150 items) and the need for a trained administrator.

Finally, the Structured Interview of Reported Systems (SIRS) has demonstrated some encouraging result in a variety of populations.[17] However, like the MS, it requires an extended administration time and a trained examiner.

The purpose of the present study was to develop and validate a brief screening

instrument, the Structured Interview Inventory of Malingered Symptomatology (SIMS). This device was intended to provide a brief screening measure in an easily administered format that could produce high rates of detection. As a screening device, the SIMS could then be used to reduce the number of individuals required to undergo more extensive assessment such as the SIRS or MS. In addition, using the SIMS as part of a battery of tests would provide confirmatory evidence of malingering rather than relying on a single instrument to make the final diagnosis.

Method

Development of the SIMS The SIMS was designed to detect malingering of five different conditions: low intelligence (LI), affective disorders (AF), neurologic impairment (N), psychosis (P), and amnestic disorders (AM). These conditions were selected, following a review of the existing literature, as some of the more commonly feigned disorders. True/false items were generated by the first author to reflect these conditions. Some items were based on questions appearing in other instruments (e.g., the MMPI) that had demonstrated some utility in detecting malingering in past research. Selected items were revised to improve their discriminability. The remaining SIMS items utilized features of malingerers noted in previous research.[18-20] Nine experienced clinical psychologists rated an initial pool of 200 items on each of the five scales as to whether each item corresponded to a particular malingered condition. A crite-

rion of 67 percent agreement among the raters resulted in a final set of 75 items (15 items per scale). This final version of the SIMS incorporated a variety of strategies to detect malingering, including the endorsement of bizarre symptoms (e.g., "Flowers have magical powers like the ability to talk to people" from the P scale), unlikely complaints (e.g., "Sometimes I lose all feeling in my hand so that it is as if I have a glove on" from the N scale), and approximate answers (e.g., "The capital of Italy is Hungary" from the LI scale). Interrater reliabilities for the final set of scales ranged from .76 (N) to .95 (AF), with a mean reliability of .84. These data supported the construct validity of the SIMS.

Subjects Subjects were undergraduate psychology students (N = 476) who volunteered for the study and received extra credit for their participation. The sample, predominately Caucasian (90%) and female (71%), was divided randomly into developmental (to determine cutoff scores) and validation groups. Subjects within each group were randomly assigned to one of eight conditions, seven of them simulating malingering conditions and one control condition. Each simulating group corresponded to one of the SIMS subscales, except for the AF scale, which was divided into those simulating mania and depression. In addition, there was a faking bad group (general malingering) and a control group that responded honestly. Subjects in the seven simulation conditions were provided with a test vignette. The vignette asked the subjects to imagine themselves detained on a serious assault charge and to simu-

late a specific disorder to avoid a serious sentence. Subjects were also warned in the vignette to "be very careful to endorse *only* the symptoms you believe will convince someone that you are suffering from the condition," because "endorsing symptoms of other disorders could result in your being caught." Pilot testing indicated that these instructions were comprehensible to the subjects and that subjects felt that they were able to malinger the conditions.

All subjects completed the SIMS, along with the F and K scales of the MMPI, the FB scale of the 16PF, and those portions of the MS that could be administered in a paper and pencil format. In its original form, many of the MS items, organized on the MgI and MgR scales, were administered orally, or the subjects were asked to respond to open-ended questions. For purposes of this study, all answers were presented in a multiple-choice format incorporating the author's scoreable options and utilizing the original scoring system.

Cutoff scores on each scale and the total score on the SIMS, F-K, FB, and the MgR and MgI scales were determined using the developmental sample. In each case, the relevant malingering group was contrasted with the control group, and cutoff scores on the various scales were selected to optimally separate the groups (i.e., detecting the most malingerers while minimizing the number of control group members incorrectly classified). These cutoff scores were then applied to the validation group to assess the discriminative power of the measures.

Results

Initial analysis of the developmental sample's responses demonstrated acceptable levels of internal consistency of the SIMS scales with reliabilities ranging from .80 to .88. Intercorrelations among the SIMS scales were generally low (e.g., .19 between the AF and LI scales), with the exception of the P and N scales (r = .75). An exploratory factor analysis confirmed generally that items of the SIMS loaded on the appropriate factor, the exception being items for the P and N scales, which tended to load on the same factor. This effect, combined with the high intercorrelation, suggested that a combination of the two scales might improve the detection rates for simulators. Consequently, the scales were combined for analyses using the validation sample. Lastly, cutoff values for each of the SIMS subscales, the total score of the SIMS, and the other validity indices were established, which maximally discriminated simulators from control subjects. These cutoff values were then applied to the validation group.

Results of the MANOVA revealed significant overall differences over all scales in all conditions (F = 11.02, df = 7, 230, $p < .01$). Univariate ANOVAs were likewise significant for each individual scale in terms of their discriminative power, with all p levels less than .0001. This suggested that each of the indices was able to significantly differentiate simulating subjects from controls.

Mean comparison procedures (Duncan's multiple range test) were used to more closely examine the response pat-

<div align="center">

Table 1
Classification Rates for Individual Scales: All Groups Combined

</div>

Scale	Sensitivity %	Specificity %	Efficiency %	χ^2
Psychosis (P)	79.5	72.7	78.6	38.705*
Amnestic disorders (AM)	88.3	90.9	88.7	101.649*
Neurologic impairment (N)	85.4	75.8	84.0	59.765*
Affective disorders (AF)	74.6	72.7	74.4	29.334*
Low intelligence (LI)	85.4	51.5	80.7	24.396*
Total score	95.6	87.9	94.5	147.668*
F	94.6	81.8	92.9	123.826*
F-K	91.2	69.7	88.2	73.971*
Faking bad (FB)	82.4	69.7	80.7	41.440*
Malingering retardation (MgR)	73.7	84.8	75.2	43.085*
Malingering insanity (MgI)	79.0	72.7	78.1	37.638*

*$p < .001$.

terns of the seven simulating conditions with the control group for each of the SIMS scales, the total score on the SIMS, F, F-K, 16PF FB, MgR, and MgI scales. These analyses indicated that all of the simulating conditions were significantly different (in the appropriate direction) from the control group ($p < .001$) using the SIMS scales and total score. Examining the other validity indices, significant difference were also found comparing the malingering groups with the control group, with the following exceptions: controls did not differ from those malingering psychosis on the F scale; controls did not differ from those malingering depression on the MgR scale; controls did not differ from those malingering neurologic impairment on the MgI scale. These analyses offered statistical evidence of the ability of the SIMS, as well as most of the other validity scales, to detect differences between honest and malingering specific conditions.

However, the major analysis of interest involved combining all conditions into a single group and to determine the rate of correct classification by each scale. This would more closely mimic conditions found in the "real" world (e.g., group membership is unknown) and would establish the heuristic value of each scale. The classification rates for each of the scales were determined using measures of sensitivity, specificity, and efficiency. Sensitivity refers to the percentage of malingerers showing the positive results on the test measure. Specificity is the percentage of honestly responding individuals who did not meet criteria values on malingering indices. The efficiency score is the percentage of subjects correctly classified into either the malingering or nonmalingering category. Results are presented in Table 1. All of the scales significantly discriminated between the control group and the malingerers. The Total SIMS score not only discriminated malingerers from control subjects to a statistically significant degree, but it also demonstrated the highest sensitivity and efficiency rates of any of the scales. It is

important to note that the cutoff scores used for the other validity indices (i.e., F, F-K, and FB) were established using the developmental sample. This maximized their ability to classify subjects within the sample. When the commonly accepted cutoff values for these scales were used,[9, 21] the efficiency rates were considerably lower (68%, 73%, and 60%, respectively).

Finally, a series of discriminant function analyses was conducted using the developmental sample to generate discriminant functions based on unit-weighted values that were then applied to the validation sample. The results indicated that the total score of the SIMS achieved rates of classification identical to those obtained by combining all of the other validity scales in predicting the presence of malingering (97.06%).

Discussion

The results of the present study suggest that the Total SIMS score was the best overall indicator of malingering. Although not able to distinguish malingerers more efficiently than the Total SIMS score, the individual SIMS scales did provided qualitative information regarding the type of symptoms the person attempted to feign (e.g., high on the AM scale indicated more malingering of amnestic symptoms).

Among the other validity scales, the F scale (MMPI) was the most effective indicator of malingering but fell slightly below that of the Total SIMS score. This efficiency rating was obtained, however, by lowering the F scale cutoff from +27, as proposed by Graham,[21] to +10. It is important to note that at +27, the F scale possessed considerably less efficiency (67.65%) but was able to correctly identify 100% of the control subjects.

In light of this finding, a procedure for detecting the presence of malingering is suggested. First, subjects who have obtained a score higher than 14 on the Total SIMS score but not less than 26 on the F scale of the MMPI, might be *suspected* of malingering and a fuller evaluation initiated. Second, only after a thorough assessment (e.g., administration of the SIRS) should the diagnosis of malingering be offered. If at any time one of these criteria is not met, the diagnosis of malingering should not be made. This strategy will enhance accuracy of detection, reduce the number and cost of false positives, provide multiple indicators of malingering, lessen the number of subjects having to undergo more extensive evaluation procedures like the SIRS, and reduce the need for trained examiners.

Although the present study demonstrated that the Total SIMS provided the highest classification rate for the presence of malingering, there are a number of limitations with this test. The greatest limitation is that this investigation is analogue-type research with limited generalizability. No criterion groups were used (e.g., subjects suffering from genuine psychosis). The incentive provided subjects to malinger (i.e., extra credit) was certainly not as compelling as that encountered by those malingerers in forensic settings (i.e., escaping criminal prosecution). Subjects were not formally screened for psychopathology. The pos-

sible impact of ethnicity or gender difference also was not explored.

One final limitation of the current study involves the issue of high base rates. Seven of eight groups, or 87.5 percent of subjects, were simulators. This rate is presumably higher than would be encountered in a real world situation. For example, rates of malingering among forensic evaluations have been reported to be 15.7 percent.[22] As a result, it could be argued that if the base rate had been lower in the present study, the efficiency of the SIMS would have been decreased. Future research is recommended to explore the efficiency ratings of the SIMS scales as well as the total score on the SIMS with varying base rates of malingering.

Despite these limitations, the present research is a critical step toward the development of a screening measure for malingering.

References

1. Hathaway SR, McKinley JC: Manual for the Minnesota Multiphasic Personality Inventory. New York: The Psychological Corporation, 1983
2. Greene RL: Assessment of malingering and defensiveness by objective personality measures, in Clinical Assessment of Malingering and Deception. Edited by Rogers R. New York: Guilford, 1988
3. Gynther MD, Altman H, Warbin R: Interpretation of uninterpretable Minnesota Multiphasic Personality Inventory profiles. J Consult Clin Psychol 40:78–83, 1973
4. Roman DD, Tuley MR, Villanueva MR, Mitchell WE: Evaluating MMPI validity in a forensic psychiatric population: distinguishing between malingering and genuine psychopathology. Crim Just Behav 17:186–98, 1990
5. Bagby RM, Rogers R, Buis T: Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. J Pers Assess 62:191–203, 1994
6. Berry DTR, Baer RA, Harris MJ: Detection of malingering on the MMPI: a meta-analysis. Clin Psychol Rev 11:585–98, 1991
7. Schretlen D, Wilkins SS, Van Gorp WG, Bobholz JH: Cross-validation of a psychological test battery to detect faked insanity. Psychol Assess 4:77–83, 1992
8. Winder P, O'Dell JW, Karson S: New motivational distortion scales for the 16PF. J Pers Assess 39:532–7, 1975
9. Krug SE: Further evidence on Sixteen PF distortion scales. J Consult Clin Psychol 47:277–84, 1978
10. Beaber RJ, Marston A, Michelli J, Mills MJ: A brief test for measuring malingering in schizophrenic individuals. Am J Psychiatry 142:1478–81, 1985
11. Gillis JR, Rogers R, Bagby RM: Validity of the M Test: simulation-design and natural-group approaches. J Pers Assess 57:130–40, 1991
12. Smith GP, Borum R: Detection of malingering in a forensic sample: a study of the M Test. J Psychiatry Law 20:505–14, 1992
13. Hankins GC, Barnard GW, Robbins ML: The validity of the M Test in a residential forensic facility. Bull Am Acad Psychiatry Law 21:111–21, 1993
14. Rogers R, Bagby RM, Gillis JR: Improvements in the M Test as a screening measure for malingering. Bull Am Acad Psychiatry Law 20:101–4, 1992
15. Smith GP, Borum R, Schinka J: Rule-out and rule-in scales for the M Test for malingering: a cross-validation. Bull Am Acad Psychiatry Law 21:107–10, 1993
16. Schretlen D: Malingering: Use of a Psychological Test Battery to Detect Two Kinds of Simulation. Ann Arbor, MI: University Microfilms International, 1986
17. Rogers R, Gillis JR, Bagby RM, Monteiro E: Detection of malingering on the Structured Interview of Reported Symptoms (SIRS): a study of coached and uncoached simulators. J Consult Clin Psychol 3:673–7, 1991
18. Resnick P: The detection of malingered mental illness. Behav Sci Law 2:21–38, 1984
19. Rogers R: Clinical Assessment of Malingering and Deception. New York: Guilford, 1988
20. Seamons DT, Howell RJ, Carlisle AI, Roe AI: Rorschach simulation of mental illness and normality. J Pers Assess 45:130–5, 1981
21. Graham JR: The MMPI: A Practical Guide (ed 2). New York: Oxford Press, 1987
22. Rogers R, Sewell KW, Goldstein AM: Explanatory models of malingering: a prototypical analysis. Law Hum Behav 18:543–52, 1994