

# Review of the Structured Inventory of Reported Symptoms-2 (SIRS-2)

**Author: Steve Rubenzer**

Steve Rubenzer, Houston, TX. Email: [rubenzer.steve@att.net](mailto:rubenzer.steve@att.net)

## Abstract

The Structured Inventory of Reported Symptoms-2 (SIRS-2) contains significant changes designed to prevent false-positive and false-negative classification errors. While the SIRS-2 has many laudatory features, the manual contains some erroneous and questionable statistics and arguments, and authors sometimes stray from the best practices advocated by the first author. The SIRS-2 is a strong choice for assessing feigned psychosis and severe psychopathology. However, evidence for its value in assessing many other conditions, particularly somatic complaints and feigned cognitive impairment, is quite limited.

**Keywords:** SIRS-2, malingering, review

---

The very first sentence of the SIRS-2 manual boldly states: “The Structured Interview of Reported Symptoms (SIRS) has been recognized in the last decade as the premier measure for the assessment of feigned mental disorders” (Rogers, Sewell, & Gillard, 2010). This review will critically examine this and other claims and attempt to objectively assess the SIRS-2's strengths and weaknesses.

## Description

The SIRS-2 is a significant revision of the original instrument, although the items and norms are virtually unchanged. Like the first version, it is a highly structured interview that assesses response styles, with a heavy emphasis on feigning or exaggeration. The SIRS-2 is administered from an examination booklet that contains instructions and the test questions, and provides for scoring of responses. The first page of the form contains a profile that allows plotting of scores, while the inner front cover contains a new decision tree for classification of response styles.

## The Revision

The SIRS revision appears to have been prompted by findings that some psychiatric populations produced much higher than expected false-positive rates (9-35%) on the original SIRS (Brand, McNary, Loewenstein, Kolos, & Barr, 2006; Rogers, Payne, Correa, Gillard, & Ross, 2009). The original SIRS manual (Rogers, Bagby, & Dickens, 1992) claimed a very low false-positive rate for various decision rules (0.5-5%), so the observed values represent a 7- to 70-fold increase in this most critical of errors. To help remedy the problem, a sample of 206 patients from the Timberlawn Mental Health System were added to 314 subjects from the original sample. Readers are referred to the

Rubenzer, S. (2010). Review of the Structured Interview of Reported Symptoms-2 (SIRS-2). *Open Access Journal of Forensic Psychology*, 2, 273-286.

original SIRS manual for information about the original sample, so new users of the SIRS-2 seem to be obligated to buy the old manual to access this important information. Timberlawn patients were described as “multiply traumatized inpatients that manifested an array of trauma, dissociative, psychotic and mood symptoms” (p. 37). They thus serve as a challenging group from which to discern those that feign psychiatric symptoms. In addition to the primary validity sample, the manual cites a standard data set from multiple clinical groups composed of 2,298 SIRS-2 protocols, of which 2,131 were administered under standard instructions and 167 given simulation instructions. Despite the new data sources, only 36 actual malingerers (as opposed to subjects asked to simulate) appear to have been used in setting the SIRS-2 decision rules (Frederick, 2010). It is important to note that the standard data set includes 647 personal-injury and disability claimants. Although a fair portion of the SIRS/SIRS-2 literature consists of work by the authors, the manual notes the SIRS has been the subject of 40 refereed articles and dissertations.

Two new indices were developed to help reduce false-positive classification errors. Because they rely on some new scales, I will defer discussion on them at this point. The SIRS-2 now has a brief scale (*Improbable Failure: IF*) to assess feigning of cognitive dysfunction, consisting of the same 20 items as in the original SIRS, although the SIRS-2 manual treats them as four items with five components each. The authors refer to “extensive validation” of the IF scale, but no references for this claim are provided. In addition, the authors note that the scale's usefulness may be limited to “persons who do not have impaired intellectual functioning” (p. 12). Subjects with an IQ below 80 were reported to make substantially more errors than those above 80. While non-psychotic disorders (anxiety, depression, PTSD) are reported to have limited effect on IF scores, there is no discussion of the impact of psychosis, use of toxic drugs, such as “wet” (marijuana dipped in formaldehyde), or traumatic brain damage.

Another new scale is the *Rare Symptom Total (RS-Total)* scale, empirically constructed of 20 items chosen to maximally differentiate between feigned and atypical but genuine clinical presentations. Only items not keyed on the eight primary scales were considered, and those chosen had very low rates of endorsements among presumed genuine clinical patients. The four IF items contribute to this scale, and have a substantial influence since their possible scores vary from 0-5 rather than 0-2, as for other items.

Two previous SIRS scales have been omitted from the revision. The *Sudden Onset (SO)* scale, which consisted of only two items, was dropped because of low reliability. The *SIRS Total score*, used for indeterminate protocols, has been replaced with the *Modified Total Index (MTI)*, described below.

As in the past, there are two primary rules for determining feigning: 1) If any of the SIRS-2's primary scales are in the *Definite feigning* range, or 2) if any three primary scales are in the *Probable feigning* range. However, two new hurdles are added. If the primary rules are met, the Rare Symptoms Total score is examined. If it exceeds a raw score of four, a classification of *Feigning* is assigned. If not, or if only one or two of the primary SIRS-2 scale are in the *Probable feigning* range, the Modified Total Index (MTI)

is examined. The MTI is the sum of four SIRS-2 primary scales, chosen for their large effect sizes in separating genuine and feigned protocols and few false-positive errors. If the MTI score is high enough, *feigning* is indicated. If moderate or low, an *indeterminate* classification is given. If very low, the new SS index is examined.

The *SS index* is designed to detect when an individual, in an attempt to avoid detection of feigning, denies nearly every symptom asked about. The SS index contains four of the SIRS-2 scales (*IF*, *Defensive Symptoms*, *Overly Specific Symptoms*, and *Direct Appraisal of Honesty*), only one of which has any obvious relationship to the described intent of the index. No data are presented to justify the scale composition or report on its internal consistency or incremental validity over the *Defensive Symptoms* scale.

A third major change in the SIRS-2 reported by the authors is an updating of reliability and validity studies. The manual contains a comprehensive appendix of SIRS studies to date, with numerous tables throughout the manual reporting reliability and validity data for various scales, indices, and populations.

### **Administration**

There are several new specifications about administration or use of the SIRS-2. The authors “strongly recommend” that one or more structured measures be administered before the SIRS-2 to discourage long, digressive answers that would be difficult to score. A special instruction is provided for subjects who continue to show verbosity. The authors also discuss use of an interpreter and plainly state, “Use of the SIRS-2 with a translator is unauthorized” (p. 15). A more troubling change is the instruction that the SIRS-2 protocol “should only be released to qualified mental health professionals with documented competency regarding SIRS-2 administration and interpretation” (p. 13). This author wonders how this is to be accomplished.

The SIRS-2 advises that new users spend two to three hours becoming familiar with the test and its questions, and the authors recommend several practice administrations supervised by a competent user. Unlike many forensic instruments, the SIRS-2 is intuitive and straightforward once the user is familiar with the test. This is an underappreciated feature, as many users may only use a specialized forensic instrument a few times a year. A test with complicated procedures can result in a botched administration or inaccurate scoring.

### **Application**

There has been concern recently that response-style indicators may produce unacceptable levels of false positives among examinees with mental retardation or borderline intellectual functioning. The SIRS-2 manual states, “The SIRS was validated primary on individuals with normal or borderline intellectual functioning. However, the SIRS-2 has also been used successfully with examinees diagnosed with mild mental retardation. Deciding whether this SIRS-2 is appropriate to use with a specific examinee must be made on a case-by-case basis” (p. 14). Yet, in one of the case examples, the authors

stated that the MMPI-2 could not be used in cases with possible intellectual limitations because it had not been validated for feigning on persons with mental retardation. The claim that the SIRS-2 has been “used successfully” is no substitute for validity data, which is much needed for an intellectually compromised population, a group that appears especially prone to false positives on a variety of validity tests (Graue, et al., 2007; Hurley & Deal, 2006; Shandera et al., 2010). In some settings, many of those who feign psychopathology will also present as cognitively impaired, and school records will not be available. In such cases, the examiner may have difficulty justifying use of the SIRS-2 if challenged, given that there is little evidence it can distinguish genuine from feigned cognitive impairment.

### **Conceptual, Technical, and Miscellaneous Considerations**

Rogers has been a pioneer in the evaluation and development of response-style measures and a leading theoretician in the design of such studies (Rogers, 1988, 1997, 2008). In the SIRS-2 manual, he and his coauthors continue to emphasize methodological issues such as the need for well-defined criterion groups and manipulation checks on simulation designs. He advocates for known-groups designs, where the criterion groups (feigning vs. honest patients) are defined using the best available measures/procedures. Yet, the criterion by which the SIRS-2 is validated is minimally described. It appears to be the clinical judgment of the treating clinician or treatment team (not specified), but there is no description of what information or processes contributed to such decisions. Did the clinician or team have access to test results like the MMPI-2, Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001) or Test of Memory Malingering (TOMM; Tombaugh, 1996)? Did they rely on observations by unit nursing or security staff? Was there any attempt to observe patients covertly? It is not even clear if clinicians had access to the SIRS-2 results, which of course, would be a fatal confound.

Rogers (1997, 2008) has frequently denounced differential prevalence designs, where different groups of subjects (outpatients vs. disability claimants) are presumed to contain different portions of feigning. The SIRS-2 manual pronounces such studies to be “simplistic and fatally flawed” (p. 73). Yet, the SIRS-2 manual repeatedly makes reference to “presumed genuine” patients, and there is no indication that these groups were screened for feigning or exaggeration. This is true of studies cited on the MMPI-2 and of the Timberlawn patients added to the primary data set. Similarly, statistics such as average IQ scores are reported for groups such as forensic examinees, with no indication whether cognitive validity tests were administered to assess if the scores were valid. This is despite an admonition early in the manual that “the domain of feigned mental disorders (including feigned psychopathology) must be considered in all clinical settings” (p. 8).

There are other areas of seeming inconsistency. The SIRS-2 now includes cognitive items in scoring of the RS-T scale. Yet the authors warn of “ill-conceived attempts to extrapolate” from different domains of feigning, such as feigning of psychopathology and cognitive impairment. The utilization of cognitive validity tests might lead to a very

different understanding of the high false positives reported on the SIRS in some samples. Would high SIRS scores in dissociative patients, for example, be viewed as false positives if sizable numbers also failed the TOMM or Word Memory Test (Green, 2005)?

The authors emphasize the negative consequences of concluding that an examinee is feigning or malingering, and examiners are instructed to adopt conservative scoring of items. However, at times the authors seem to overreach in finding alternatives to a malingering explanation. In one example (Mr. Simon), one or more of the authors were apparently retained to provide a second opinion on a defendant. In the first evaluation, the defendant reportedly endorsed every item (raw score = 25) on the M-FAST (recommended cutoff scores for screening purposes is greater or equal to six) and scored 17 and 18 on Trial 2 and the Retention Trial of the TOMM, respectively. The first TOMM score is below chance at the .05 level, which is widely regarded as definitive evidence of negative response bias (Berry & Schipper, 2008; Boone, 2007; Slick, Sherman, & Iverson, 1999; Sweet, Condit, & Nelson, 2008). Combining the scores from the two trials yields 35 correct out of 100, which is likely to occur less than one time in a thousand by chance. Yet, the authors attribute this performance to “distractibility,” a factor that has not been demonstrated to seriously affect TOMM or other cognitive validity test scores in the absence of psychosis (Goldberg, Back-Madruga, & Boone, 2007; Sollman, Ranseen, & Berry, 2010), and certainly cannot explain below-chance performance.

The authors discuss at length issues involved in labeling response distortion, and caution about applying the label of “malingering” based on DSM-IV-TR criteria or from test results. A clear distinction is made between malingering and factitious motivations, with recommendations on distinguishing the two using the SIRS-2, Personality Assessment Inventory (Morey, 1991), and specific interview questions. No discussion is provided about the SIRS-2's applicability or sensitivity to the “unconscious” exaggeration sometimes said to occur in somatoform disorders.

### **Psychometrics**

The manual provides a detailed description of expected statistics, typically broken down by scale and summary decision. For most, both correlation and concordance of agreement statistics are provided—an important matter when agreement on actual scores between raters counts. Internal and test-retest reliabilities are generally adequate for the individual scales, although about 20% fall below the minimum standard (.80) advocated by Heilbrun (1992). Interrater reliabilities were quite high, typically exceeding .98. However, several caveats are in order. These figures were obtained within a formal research program, where raters were trained and supervised by people closely associated with the instrument. It is unclear if such figures could be obtained in more typical settings. In addition, interrater agreement was assessed through Pearson's  $r$ , which is sensitive to the rank order agreement between two raters, but not difference in mean values. In other words, one rater could consistently assign higher scores than another, but as long as subjects were ranked the same by the two raters, the Pearson's  $r$  would approach 1.00. For this reason, Shrout and Fleiss (1979) recommended use of the

intraclass coefficient to capture both sources of variance. Despite these issues, the SIRS-2's scoring is objective and interrater reliability is likely to be less problematic than for most forensic assessment instruments.

The authors also report standard errors of measurement (SEM) for each SIRS-2 scale using *interrater* reliability figures. In the first SIRS manual, internal consistency coefficients were used for calculating the SEM. In a footnote, the authors state, "The proper focus is on examinees and the independent accuracy of measuring their scores rather than the homogeneity of SIRS items." No further support for this statement is given, and the net effect is to produce very small estimates of the SEM for most scales. The authors then compare these with those calculated for MMPI-2 validity scales using internal consistency as the reliability figure. It seems internal consistency is only relevant to error estimates for self-report measures, not structured interviews. If SEMs are to be meaningful, they should arguably take account of all sources of variance (interrater, internal consistency, and test-retest) and a generalizability coefficient would serve as the basis for estimation (Suen, 1990). Estimations based on only one source of error, especially when one is much smaller than the others, are likely to be overly conservative and misleading. However, it should be noted that the internal reliability figures for SIRS-2 scales are considerably higher than those for the MMPI-2 or RF Fp/Fp-r and the RF Fs scales.

Validity evidence is presented for each of the SIRS-2 scales in the form of correlations with MMPI and MMPI-2 response-style scales, including F, Fb, Fp, F-K and Ds/Dsr, as well as validity scales from the Personality Assessment Inventory (Morey, 1991), M-FAST, and the Structured Inventory for Malingered Symptomology (Widows & Smith, 2005). Data from MMPI and MMPI-2 are combined and relative numbers of each not reported. No data are presented for newer generation MMPI-2/MMPI-2-RF response scales like Fs (Ben-Porath & Tellegan, 2008), FBS/FBS-r (Lees-Haley, 1992; Ben-Porath & Tellegan, 2008), the Henry Heilbronner Index (Henry, Heilbronner, Mittenberg, & Enders, 2006), Meyer's Validity Index (Meyers, Millis, & Volkert, 2002), or Response Bias Scale (Gervais, Ben-Porath, Wygant, & Green, 2007), which are designed to detect feigning of somatic complaints or corroborate feigning of cognitive impairment. Most correlations of MMPI and other instruments' response scales with SIRS/SIRS-2 scales are large and show expected patterns. The manual reports correlations for RS-Total, which are generally lower than the SIRS-2 primary scales, but not for MT Index.

Several other forms of validity data are presented. Cohen's *ds* of the SIRS-2 primary scales are reported to average about 1.45 with simulators versus clinical samples, excluding two studies of primarily dissociative-disorder patients. Weighted average *ds* for known group and bootstrapping designs range from 1.50 to 2.50. It should be noted that two of the four studies cited in this area are by Rogers, and one is a dissertation. A separate section reports that the MT Index had a Cohen's *d* of 1.85, but does not report the study design. The manual references a published factor analysis of SIRS scales (Rogers, Jackson, Sewell, & Salekin, 2005) in which a two-factor solution (*Unlikely detection scales* and *Amplified detection scales*) was found.

The authors present evidence of discriminate validity by correlating the primary SIRS-2 scales with the MMPI-2 K scale. While low, negative correlations were observed, consistent with expectations, this is not a demanding or particularly meaningful test. More informative would be an examination of whether SIRS-2 scales correlate substantially with MMPI-2 *clinical* scales in a patient sample independently assessed and culled for feigning. While an optimal result would be no substantial correlation, indicating the response-style scales are not affected by actual psychopathology, it is likely a substantial positive correlation would be observed. In fact, this appears to be the case: The Clinical-General composite normative group has mean values on *Subtle Symptoms*, *Selectivity*, and *Severity of Symptoms* scales that are in the upper end of the *Indeterminate* range, and the score on *Subtle Symptoms* is virtually the same as the malingering group. The inclusion of the Timberlawn patients, who presented as highly symptomatic, may be responsible for this finding. More on this will be presented later.

Generalization of validity across ethnicities was examined in three studies. No difference was observed between Anglos and African-Americans, while the SIRS-2 scales correlated with MMPI-2 validity scales best for Hispanics. Discriminate validity is examined by ethnicity and gender, and while it appears comparable across groups, no results of statistical tests are reported. Here and in the tables in the validity section, the number of subjects contributing to the analyses is not reported.

### **Diagnostic Statistics**

Diagnostic statistics are widely used in medicine, where diagnoses are viewed as natural categories, and a disease is present or not. Many diagnostic tests are interpreted in dichotomous fashion, with a positive result indicating presence of a condition or disease, a negative finding with its absence. Thus there are four possible combinations, as illustrated in the fourfold classification table (see Figure 1).

**Figure 1**

Fourfold Classification Table

		SIRS-2 Classification	
		Feign	Honest
Criterion	Feign	True Positives	False Negatives
	Honest	False positives	True Negatives

Note. Labels in green font represent accurate decisions, those in red denote errors.

In calculating diagnostic statistics, the authors did not include subjects who could not be reliably classified on the SIRS-2 as honest or feigning. The number of indeterminate cases was substantial (120 of 522), and there have been posts to professional listservs questioning whether figures such as sensitivity and specificity are legitimately calculated when there are indeterminate cases. For example, sensitivity is often defined as the probability that a test will detect a condition when it is present. Sensitivity would be calculated by dividing the number of correctly identified feigning cases by the total number of feigners. Leaving out the indeterminate cases will overestimate the sensitivity. With so many indeterminate cases, any estimate of diagnostic statistics will be subject to a large range of uncertainty.

The sensitivity of the SIRS-2 decision rules is reported to be .80, with a false-positive rate of 2.5%. However, the numbers given by the authors (10 false positives out of 402 classified subjects) indicate the false-positive rate was calculated incorrectly, as the denominator should be the total number of *honest* subjects, not the grand total. Figured this way, the false-positive rate is 3.6%. The data in Figure 2 is not directly reported in the manual, but was recreated from information in it and, with the exception of the false-positive rate, produces diagnostic statistics that match those presented by the authors.



**Figure 2**

Diagnostic Table for SIRS-2 Sample Excluding Unclassified Cases

		SIRS-2 Classification		Total
		Feign	Honest	
Criterion	Feign	102	26	128
	Honest	10	264	274
	Total	112	290	402

The reported sensitivity and false-positive rates for the SIRS-2 are quite different from the original version. The original SIRS was reported to have a sensitivity of .485, meaning less than half (48.5%) of those feigning would be so classified. The SIRS-2 manual reports a sensitivity of .80, despite implementing two new decision rules that restrict assignment of a feigning classification. Frederick (2010) showed that this difference can be attributed to the exclusion of unclassified subjects in the SIRS-2 sample: By excluding the harder cases, the *apparent* sensitivity is greatly increased. The original SIRS manual did not provide a clear estimate of the false-positive rate using all the decision rules. Rather, individual scales were reported to have false-positive rates from 0-5% at the *Definite feigning* level. The rule for three or more scales in the *Probable feigning* range was reported to have a false-positive rate of 0.5%. The false-positive rate for the SIRS-2, figured from the data above, appears higher than the previous version, despite addition of the new decision rules. This is almost certainly due to the inclusion of the highly symptomatic Timberlawn patients in the validation group.

The authors include a novel discussion of base rates and their use in interpreting SIRS-2 results, and come to a startling conclusion: "Overall prevalence rates of malingering are inapplicable to the SIRS-2. . . . Applying local base rates decreases rather than increases the accuracy of utility estimates" (p. 39). No data are presented in support of this assertion. The authors argue that base rates vary dramatically among forensic settings, yet also note that the SIRS-2 will be typically administered to only a select group of examinees for whom malingering is suspected. Among such a select group, the base rate may easily be .60 or more, and the range of values across settings is likely restricted. The authors do give tentative values for *Positive Predictive Power* (PPP; the percentage of those failing the SIRS-2 who are feigning) and *Negative Predictive Power* (NPP; the percentage of those passing the SIRS-2 who are indeed genuine patients) based on the base rate in their sample of 31.8%. Both were .91. However, this figure will significantly underestimate the PPP and overestimate the NPP if higher base rates

are present or brought about by preselection of cases. For example, Jackson, Rogers, and Sewell (2005) found that M-FAST scores equal to or greater than six produced a PPP of .74 in predicting SIRS classification of feigning. If we accept SIRS failure as equivalent to feigning, this figure (.74) is effectively a base rate for the SIRS assessment brought about by use of a screening instrument.

### **Conclusion**

The SIRS-2 has the same strengths as its predecessor: standardized administration and scoring, high interrater reliability, and validation in both simulation and known group studies. The alarming false-positive rate in the original SIRS among dissociative patients provided impetus to create new checks on the classification rules, and initial data indicate these are effective in reducing false-positive errors. However, The SIRS-2 decision rules have been validated on a relatively small group of actual malingerers. Although the Timberlawn patients provide a challenging group for a feigning measure, reliance on any particular sample is risky. How well will the new MTI and RS-Total rules work in other settings or other severe populations? Cross validation by independent researchers is an essential aspect of scientific inquiry, and is doubly important when the findings can greatly impact individual lives. The original SIRS was thought to have very low false-positive rates before being applied to the Timberlawn patients, and it is likely that many users cited this belief when discussing SIRS results or in reports or courtroom testimony. While it appears the authors have moved quickly to address this disturbing finding, the stage is set for a second act should similar results be found among the borderline retarded, for example. Conversely, I wonder if significant numbers of the Timberlawn patients were not dramatizing or over reporting, perhaps as a factitious presentation, and if the SIRS revision might have been altogether unnecessary. With only the data presented, it's not possible to say.

The manual does not clearly present information that is important to evaluate the SIRS-2's performance. Among these are the procedures used to determine if subjects were malingering or honest, and a complete tabulation of subjects by SIRS-2 results and criterion. Outside of psychology, there is a concerted effort to standardize reporting of diagnostic studies. The STARD initiative (Bossuyt, et al., 2003) specifies 25 checklist items that researchers should address and report. Most have to do with clearly and completely describing selection of subjects, instruments, procedures, and data. More than 200 scientific journals in a variety of medical and scientific fields have now endorsed the STARD standards or referenced them. While the guidelines are intended to apply to individual studies, such information also belongs in test manuals.

While the SIRS-2 has a claim as the instrument of choice for malingering of major psychiatric disorders, data for feigning of less severe conditions, such as anxiety, depression, PTSD (see Rubenzer, 2009), and cognitive limitations is more limited and generally limited to papers by the author or dissertations. For somatoform conditions, data are virtually non-existent. The SIRS-2 does not rival the MMPI-2 or MMPI-2-RF in the latter two areas, and the claim at the beginning of the paper seems excessively expansive. Users should also be aware that, because of conservative decision rules,

substantial numbers of feigners will not be identified on the SIRS-2, which should limit its use, without supportive data, as a criterion for validating other instruments.

The presence of significant errors, omissions, and biased comparisons in the SIRS-2 manual suggests forensic tests should undergo peer review *prior to publication*. Authors are unlikely to be completely objective about their own instrument, and they may develop blind spots or biases pertaining to a particular assessment topic. There may also be a tendency to promote or oversell the instrument at the expense of competitors or alternative approaches. Instruments are vital to forensic psychology as a profession, and the SIRS-2 holds a special place because of the importance of response-style assessment and the stature of the first author: It serves as a flagship for the profession. It should embody the very best the field has to offer, and peer review is the accepted quality-control mechanism in academic endeavors.

Many of the SIRS-2's present shortcomings could be addressed through a revised manual. The need for cross validation by others, however, is more substantive and will likely take some time. Until it is forthcoming, users should value and employ the SIRS-2 for assessment of feigned serious psychopathology, but should avoid relying on any single test as a gold standard.

**Acknowledgements:** Prior to writing this review, some detailed and thoughtful comments on the SIRS-2 were posted on listservs by Richard Frederick, Barry Rosenfeld, Christopher Ray, and Greg DeClue. Ideas and analyses presented here undoubtedly benefited from these insightful contributions. Greg DeClue also contributed suggestions on an initial draft.

received July 3, 2010, revision submitted July 10, 2010; accepted July 30, 2010

## References

- Ben-Porath, Y., & Tellegan, A. (2008). MMPI-2 RF™ (Minnesota Multiphasic Personality Inventory-2 Restructured Form™). Pearson Assessment Systems. Minneapolis: University of Minnesota Press.
- Berry, D. T., & Schipper, L. J. (2008). Assessment of feigned cognitive impairment using standard neuropsychological tests. In R. Rogers (Ed.), *Clinical assessment of malingering and deception*, 3<sup>rd</sup> ed. New York: Guilford.
- Boone, K. B. (2007). A reconsideration of the Slick et al. criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective*. (pp. 29-49). New York: Guilford Press.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D., & de Vet, H. C. W. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clinical Chemistry*, 49, 1-6. <http://www.stard-statement.org/>
- Brand, B. L., McNary, S. W., Loewenstein, R. J., Kolos, A. C., & Barr, S. R. (2006). Assessment of genuine and simulated dissociative identity disorder on the structured interview of reported symptoms. *Journal of Trauma & Dissociation*, 7, 63-85.
- Frederick, R. (2010). Unpublished manuscript. April 3, 2010.
- Gervais, R. O., 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.
- Goldberg, H. E., Back-Madruga, C., & Boone, K. B. (2007). The impact of psychiatric disorders on cognitive symptom validity tests. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective*. (pp. 281-309). New York: Guilford Press.
- Graue, L., Berry, D., Clark, J., Sollman, M., Cardi, M., Hopkins, J., & Werline, D. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings. *Clinical Neuropsychologist*, 21(6), 929-942.
- Green, P. (2005). *Green's Word Memory Test user's manual*. Edmonton: Green's Publishing, Inc.
- Heilbrum, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16(3), 257-272.

- Henry, G. I., Heilbronner, R., Mittenberg, W., & Enders, C. (2006). The Henry-Heilbronner Index: A 15-Item empirically derived MMPI-2 subscale for identifying probable malingering in personal injury litigants and disability claimants. *The Clinical Neuropsychologist* 20(4), 786-797.
- Hurley, K. E., & Deal, W. P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues. *Mental Retardation*, 44(2), 112-119.
- Jackson, R. L., Rogers, R., & Sewell, K. W. (2005). Forensic applications of the Miller Forensic Assessment of Symptoms Test (M-FAST): Screening for feigned disorders in competency to stand trial evaluations. *Law and Human Behavior*, 29, 199-210.
- Lees-Haley, P. R. (1992). Efficacy of MMPI-2 validity scales and MCMI-II modifier scales for detecting spurious PTSD claims: F, F-K, Fake Bad Scale, ego strength, subtle-obvious subscales, DIS, and DEB. *Journal of Clinical Psychology*, 48(5), 681-689.
- Meyers, J. W., Millis, S. R., & Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology*, 17(2), 157-169.
- Morey, L. (1991). *The Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Miller, H. A. (2001). *M-Fast: Miller Forensic Assessment of Symptoms Test professional manual*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R. (1988). Introduction. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 250-268). New York, NY: Guilford Press.
- Rogers, R. (1997). Researching deception. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2<sup>nd</sup> ed., pp. 398-426). New York, NY: Guilford Press.
- Rogers, R. (2008). Researching deception. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3<sup>rd</sup> ed., pp. 411-434). New York, NY: Guilford Press.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Rogers, R., Jackson, R. L., Sewell, K. W., & Salekin, K. L. (2005). Detection strategies for malingering: A confirmatory factor analysis of the SIRS. *Criminal Justice and Behavior*, 32, 511-525.

- Rogers, R., Payne, J. W., Correa, A. A., Gillard, N. D., & Ross, C. A. (2009). A study of the SIRS with severely traumatized patients. *Journal of Personality Assessment*, 91(5), 429–438.
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms (SIRS) and professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Rubenzler, S. J. (2009). Posttraumatic Stress Disorder: Assessing response style and malingering. *Psychological Injury and Law*, 2(2), 114-142. DOI 10.1007/s12207-009-9045-4 <http://www.springerlink.com/content/5170u2j543702288/>
- Shandera, A. L., Berry, D. T. R., Clark, J. A., Schipper, L. J., Graue, L. O., & Harp, J. P. (2010). Detection of malingered mental retardation. *Psychological Assessment*, 22(1), 50–56.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(40), 545-561.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 8(2), 420-428.
- Suen, H. K. (1990). *Principles of test theories*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sollman, M. J., Ranseen, J. D., & Berry, D. T. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, 22(2), 325-335.
- Sweet, J. J., Condit, D. C., & Nelson, N. W. (2008). Feigned amnesia and memory loss. In R. Rogers (Ed.), *Clinical assessment of malingering and deception (3<sup>rd</sup> ed.)*. New York: Guilford Press.
- Tombaugh, T. N. (1996). *The Test of Memory Malingering (TOMM)*. Toronto, Canada: Multi-Health Systems.
- Widows, M., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomology (SIMS) and professional manual*. Odessa, FL: Psychological Assessment Resources.