Psychological Assessment
2013, Vol. 25, No. 4, 1367–1369

DOI: 10.1037/a0033696

BRIEF REPORT

Rates of Computational Errors for Scoring the SIRS Primary Scales

Elizabeth A. Tyner and Richard I. Frederick
U.S. Medical Center for Federal Prisoners

We entered item scores for the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1991) into a spreadsheet and compared computed scores with those hand-tallied by examiners. We found that about 35% of the tests had at least 1 scoring error. Of SIRS scale scores tallied by examiners, about 8% were incorrectly summed. When the errors were corrected, only 1 SIRS classification was reclassified in the fourfold scheme used by the SIRS. We note that mistallied scores on psychological tests are common, and we review some strategies for reducing scale score errors on the SIRS.

Keywords: malingering, test scoring errors, forensic assessment

The Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1991, and now SIRS-2; Rogers, Sewell, & Gillard, 2011) is an individually administered structured interview intended to evaluate the prospect that symptoms of psychopathology are feigned. The SIRS has 172 items, primarily designated as "Detailed Inquiries" or as "General Inquiries." Most of the items (n = 168) yield a score of 0, 1, or 2. The Primary Scales of the SIRS are scored by summing selected items. Primary Scales generated from General Inquiries (i.e., Rare Symptoms [RS], Symptom Combination [SC], Improbable or Absurd [IA], or Reported vs. Observed [RO]) comprise seven, eight, or 12 items. Primary Scales generated from Detailed Inquiries (Blatant [BL], Subtle [SU], Severity [SEV], or Selectivity [SEL]) comprise 15, 16, or 17 items.

To compute the score for RS, SC, IA, or RO requires searching the booklet for colored symbols printed next to scores and collecting those scores for summation. Computing scores for BL, SU, SEV, and SEL requires counting or summing certain endorsed items on two separate pages of detailed inquiries—some of these items are shaded, and some are unshaded. Then the sums on each page are summed and recorded on the front cover of the SIRS booklet. Rogers et al. (2011) referred to these processes as "simple tasks" (p. 29). The publisher of the SIRS (Psychological Assessment Resources) does not provide a computerized scoring method for the SIRS or SIRS-2. All scores are generated by the counting and arithmetic skills of the person who administers the test.

This article was published Online First July 15, 2013.

Elizabeth A. Tyner and Richard I. Frederick, Department of Psychology, U.S. Medical Center for Federal Prisoners.

The conclusions and opinions reported in this article do not necessarily represent the opinions of the Federal Bureau of Prisons or the U.S. Department of Justice.

Correspondence concerning this article should be addressed to Richard I. Frederick, 334 East Kearney #323, Springfield, MO 65803. E-mail: rickfrederick@gmail.com

Method

We were interested in the rate at which errors occur when scoring the SIRS. We located in case files completed and scored SIRS test protocols that had been administered at a large inpatient forensic assessment facility from March 1993 to April 2010. We first note that we collected 172 cases. During the same time period at this facility, several thousand Minnesota Multiphasic Personality Inventory-2s (MMPI-2; Butcher, Dahlstron, Graham, Tellegen, & Kaemmer, 1989) were administered. In general, at this facility, the SIRS was a follow-up assessment to the MMPI-2 and now the MMPI-2-Restructured Form (Tellegen & Ben-Porath, 2008/2011), especially when those tests indicated the presence of feigning. We did not note a single instance in which the SIRS had been administered as a stand-alone assessment.

We identified the professional level of the examiners for each of the 172 records, which included 94 by psychologists, 51 by psychology interns, 10 by postdoctoral psychology trainees, nine by psychology practicum students, and eight by one psychiatrist. To minimize potential disclosure of error rates for individuals, we evaluated rates of errors by psychologists and nonpsychologists (students, interns, postdocs, and psychiatrist; see Table 1).

We entered the scored responses to each item and the computed scores into an Excel workbook. We note that it was not possible, in general, to evaluate the accuracy of the score assigned to an examinee's response. Some examiners recorded the examinee's response to every query (yes, no, sometimes). Some merely circled the score they assigned to the item. Once the item scores were entered into Excel, we verified them and corrected any misentries. We then generated formulas in Excel to sum up the appropriate items to generate the primary scale totals. We then compared those totals with the totals computed by the examiners and identified errors of computation. When Excel indicated there had been an error in scoring, we once again checked our data entries against the values recorded in SIRS booklets until we were sure the error was from the examiner and not from our data entry.

Table 1
SIRS Record Error Rates

Number of errors	Number of records	Psychologists	Nonpsychologists
0	112 (65.1%)	63	49
1	32 (18.6%)	15	17
2	18 (10.5%)	11	7
3	5 (2.9%)	1	4
4	3 (1.7%)	3	0
6	1 (0.6%)	0	1
7	1 (0.6%)	1	0

Note. SIRS = Structured Interview of Reported Symptoms.

Results

An *error* was defined as a discrepancy between the scale totals recorded by the examiner and derived by Excel. There were a total of 108 (7.8%) errors in 1,376 (i.e., 172 × 8 primary scales) opportunities to make an error. Table 1 summarizes the number of errors by record form: 60 of 172 (34.9%) records had at least one primary scale computed score error. Almost all errors were errors of computation. A few errors were so discrepant (e.g., 17 vs. four) that they most likely were transcription errors (e.g., the total for one scale was assigned to another). There were six psychologists who administered the SIRS—all are currently board certified in forensic psychology by the American Board of Professional Psychology. All of these psychologists made at least one scoring error. In 31 of 94 occasions (33.0%), psychologists made at least one scoring error. In 29 of 78 occasions (37.2%), nonpsychologists made at least one scoring error.

Table 2 demonstrates that errors were not evenly distributed among primary scales. Except for the RS, SC, and IA scores, which are obtained by simple summing of items identified by color-coded symbols, there appears to be no great advantage for psychologists in avoiding errors.

We were interested in what the consequences of these errors were to classification of primary scale scores and to the overall classification of the SIRS performance. Primary scores are classified as Definitely Feigned, Probably Feigned, Indeterminate, or Honest. These four categories comprise a potentially broad range of scores, and so a computational error might not result in any change of primary score classification. Of 108 scoring errors, 23 scale scores (21.3%) resulted in a different classification for the primary scale when the error was corrected. Sixty records had at least one error in the computation or transcription of the SIRS primary score; 17 records had instances of primary scores that changed scale classifications when computed correctly.

Overall SIRS performances can be classified as Definitely Feigned, Probably Feigned, Indeterminate, or Honest. Of the 17 records with at least one primary scale misclassification, when the primary scales were correctly classified, only one record changed overall classification—from Probably Feigned to Definitely Feigned. Although it is tempting to view this outcome as an indication of robustness of the SIRS against overall misclassification from computational errors, in this one instance, RS should have been scored 5, but was scored 4. A miscount of 1 point for one scale changed the overall classification of the record (see Table 3).

Table 2
Rate of Computational Errors for SIRS Primary Scales

Primary scale	Errors	Psychologists	Nonpsychologists
RS	5 (4.6%)	1 (1.7%)	4 (8%)
SC	6 (5.6%)	1 (1.7%)	5 (10%)
IA	3 (2.8%)	1 (1.7%)	2 (4%)
BL	29 (26.9%)	15 (25.9%)	14 (28%)
SU	28 (25.9%)	20 (34.5%)	8 (16%)
SEL	11 (10.2%)	5 (8.6%)	6 (12%)
SEV	16 (14.8%)	9 (15.5%)	7 (14%)
RO	10 (9.3%)	6 (10.3%)	4 (8%)
Total	108	58	50

Note. There were 108 computational errors for Primary Scale score in 172 records. Eight chances for error per record. There were 1,376 chances for error. Rate of errors = 7.8%. SIRS = Structured Interview of Reported Symptoms; RS = Rare Symptoms; SC = Symptom Combinations; IA = Improbable or Absurd; BL = Blatant; SU = Subtle; SEL = Selectivity; SEV = Severity; RO = Reported versus Observed.

Discussion

We note that rates of errors seem to increase with the complexity of the scoring process. The simplest computations are for RS, SC, and IA. The rates of errors for those primary scales were about 3%–6%. Finding the values for SEL and SEV requires finding sums twice. The rates of errors for those primary scales were about 10%–15%. Finding the values for SEL and SEV involves *counting*, but finding the values for BL and SU involves *summing* and involves discriminating between shaded and nonshaded items to find the values to sum. The rates of errors for BL and SU were about 18%–27%.

RO would seem to involve the same process as finding the values for RS, SC, and IA—one takes the sum of various RO items across a number of pages. What we noted was that, unlike the items comprising RS, SC, and IA, the RO items often had two scores for a single item, a 1, and then a 2, if the queried symptom became worse. The presence of two scores likely explains the higher rate of errors for RO of about 9%.

Being aware of the susceptibility of examiners to make errors is the first step in overcoming them. We offer suggestions on how to check a SIRS protocol for potential errors. First of all, we note a simple method to immediately identify errors in sums for BL, SU, SEL, and SEV on Detailed Inquiries I and Detailed Inquiries II. It is straightforward algebraically to demonstrate that the sum of BL

Table 3
SIRS Primary Scale Classification Errors

Derived scale classification	Correct scale classification	Instances of misclassification
Honest	Indeterminate	3
Indeterminate	Probable	5
Probable	Indeterminate	3
Probable	Definite	2
Definite	Probable	6
Indeterminate	Definite	2
Definite	Indeterminate	1
Honest	Definite	1

Note. SIRS = Structured Interview of Reported Symptoms.

and SU should always equal the sum of SEL and SEV for each set of Detailed Inquiries and for the final totals. Examiners should add BL + SU and compare with SEL + SEV for each set of Detailed Inquiries. Then examiners should compare BL + SU with SEL + SEV for the total scores. Any inequalities immediately identify errors in scoring. Unfortunately, equalities do not necessarily mean the items were summed correctly.

Examiners might consider other strategies to minimize computational errors on the SIRS and SIRS-2. The most straightforward check would be to ask a colleague to review the record for errors. A more involved method might be to transfer scores from the 14-page SIRS booklet into another medium (e.g., a spreadsheet) for summing. Examiners could enter item scores by primary scale onto a single page (e.g., some scoring summary page they created for themselves) and avoid depending on their working memory to keep track of scores as they go through the booklet. It is also easy to set up an Excel workbook to enter the items scores and to instantly score the SIRS primary and secondary scales. This would be helpful, especially because the SIRS-2 has a much more complex hurdled approach to overall classification than does the SIRS. This hurdled approach to decision making is easily accomplished in Excel or other spreadsheets and precludes examiner errors in making hurdle decisions. One need only enter the items scores and verify them, and then computational errors are eliminated. It is most desirous that the SIRS publisher develop some scoring software to eliminate the problems of computational and decision errors.

We note, however, that errors in scoring psychological tests occur at every level, including transferring the items scores to another medium for scoring (Allard, Butler, Faust, & Shea, 1995; Allard & Faust, 2000; Simons, Goddard, & Patton, 2002). Allard et al. found that over 50% of the time, an objective personality contained at least one scoring error (by mistallying), and about 20% of the time, the error was sufficiently significant to alter the clinical interpretation of the scale. Allard and Faust found similar results when looking at three different tests, but they also found mistakes when items responses were keyed into computer programs. They also discovered that one scoring program was referencing the wrong norms to derive T scores. We echo the recommendations of Allard and Faust (2000) that "[Testing] sites . . . perform a fairly straightforward audit of their scoring practices to determine risk for gross errors" (p. 127).

References

Allard, G., Butler, J., Faust, D., & Shea, M. T. (1995). Errors in hand scoring objective personality tests: The case of the Personality Diagnostic Questionnaire—Revised (PDQ–R). *Professional Psychology: Research and Practice, 26,* 304–308. doi:10.1037/0735-7028.26.3.304

Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment, 7,* 119–129. doi:10.1177/107319110000700203

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2*. Minneapolis: University of Minnesota Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1991). *Structured Interview of Reported Symptoms: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., Sewell, K. W., & Gillard, N. D. (2011). *Structured Interview of Reported Symptoms* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Simons, R., Goddard, R., & Patton, W. (2002). Hand-scoring error rates in psychological testing. *Assessment, 9,* 292–300. doi:10.1177/1073191102009003008

Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form) technical manual*. Minneapolis: University of Minnesota Press.

Received December 13, 2012
Revision received May 28, 2013
Accepted May 30, 2013 ■