

Eastern Kentucky University

Encompass

---

Psychology Faculty and Staff Research

Psychology

---

3-1-2010

## Utility of the MMPI–2-RF (Restructured Form) Validity Scales in Detecting Malingering in a Criminal Forensic Setting: A Known-Groups Design

Martin Sellbom

*University of Alabama - Tuscaloosa*

Joseph A. Toomey

*John Jay College of Criminal Justice*

Dustin B. Wygant

*Eastern Kentucky University, [dustin.wygant@eku.edu](mailto:dustin.wygant@eku.edu)*

L. Thomas Kucharski

*John Jay College of Criminal Justice*

Scott Duncan

*United States Penitentiary, Atlanta, Georgia*

Follow this and additional works at: [https://encompass.eku.edu/psychology\\_fsresearch](https://encompass.eku.edu/psychology_fsresearch)



Part of the [Clinical Psychology Commons](#)

---

### Recommended Citation

Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI–2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment*, 22(1), 22–31. <https://doi.org/10.1037/a0018222>

This Article is brought to you for free and open access by the Psychology at Encompass. It has been accepted for inclusion in Psychology Faculty and Staff Research by an authorized administrator of Encompass. For more information, please contact [Linda.Sizemore@eku.edu](mailto:Linda.Sizemore@eku.edu).

# Utility of the MMPI–2-RF (Restructured Form) Validity Scales in Detecting Malingering in a Criminal Forensic Setting: A Known-Groups Design

Martin Sellbom  
The University of Alabama

Joseph A. Toomey  
John Jay College of Criminal Justice

Dustin B. Wygant  
Eastern Kentucky University

L. Thomas Kucharski  
John Jay College of Criminal Justice

Scott Duncan  
United States Penitentiary, Atlanta, Georgia

The current study examined the utility of the recently released Minnesota Multiphasic Personality Inventory—2 Restructured Form (MMPI–2-RF; Ben-Porath & Tellegen, 2008) validity scales to detect feigned psychopathology in a criminal forensic setting. We used a known-groups design with the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) as the external criterion to determine groups of probable malingering versus nonmalingering. A final sample of 125 criminal defendants, who were administered both the SIRS and the MMPI–2-RF during their evaluations, was examined. The results indicated that the two MMPI–2-RF validity scales specifically designed to detect overreported psychopathology, F-r and F<sub>p</sub>-r, best differentiated between the malingering and nonmalingering groups. These scales added incremental predictive utility to one another in this differentiation. Classification accuracy statistics substantiated the recommended cut scores in the MMPI–2-RF manual (Ben-Porath & Tellegen, 2008) in this forensic setting. Implications for these results in terms of forensic assessment and detection of malingering are discussed.

*Keywords:* MMPI–2-RF, validity scales, malingering, forensic assessment

It is vital that psychologists have instruments at their disposal that can effectively determine the veracity of symptom presentation in forensic psychological evaluations given the high stakes associated with such proceedings. It is not uncommon for criminal defendants to feign psychopathology (particularly psychosis) during the course of competency and criminal responsibility evaluations to avoid or delay punishment (Rogers, 2008). To date, the Minnesota Multiphasic Personality Inventory—2 (MMPI–2; Butcher et al., 2001) is one of the most widely used instruments of any general assessment measure (i.e., personality assessment, intelligence assessment, neuropsychological assessment, etc.) used by clinical psychologists across various clinical and forensic settings (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Camara, Nathan, & Puente, 2000; Lally, 2003). The instrument has also been touted by neuropsychologists as one of the top five

most widely used instruments in the determination of response bias (Sharland & Gfeller, 2007). Additionally, numerous meta-analyses (e.g., Berry, Baer, & Harris, 1991; Rogers, Sewell, Martin, & Vitacco, 2003; Rogers, Sewell, & Salekin, 1994) have found the MMPI–2 Infrequency scales effective in identifying overreported psychopathology.

Most of the studies included in the aforementioned meta-analyses, however, have involved analogue simulation designs. These research designs typically involve the administration of the MMPI–2 to psychologically well-adjusted college students under the instruction to feign symptoms of mental illness and comparing them to psychiatric patients (e.g., Graham, Watts, & Timbrook, 1991). Though simulation studies provide important experimental control, they are often limited with regard to external validity, making it difficult to determine the degree to which the results generalize to actual forensic evaluations. Rogers et al. (1994) and Rogers and Bender (2003) specifically noted that the majority of the malingering research involving the MMPI–2 has been void of actual malingerers.

Although they represent the standard in cognitive malingering research, studies utilizing a known-groups design (i.e., defining symptom-overreporting groups based on an empirically validated response bias criterion) are scarce with regard to feigned psychopathology and are sometimes confounded by methodological con-

---

Martin Sellbom, Department of Psychology, The University of Alabama; Joseph A. Toomey and L. Thomas Kucharski, Department of Psychology, John Jay College of Criminal Justice; Dustin B. Wygant, Department of Psychology, Eastern Kentucky University; Scott Duncan, United States Penitentiary, Atlanta, Georgia.

Correspondence concerning this article should be addressed to Martin Sellbom, Department of Psychology, 348 Gordon Palmer Hall, The University of Alabama, Tuscaloosa, AL 35487. E-mail: msellbom@ua.edu

cerns. For example, Gassen, Pietz, Spray, and Denney (2007) found Megargee's (2004) Criminal Offender Infrequency (Fc) scale produced the highest hit rate of any MMPI-2 validity scale in detecting criminal defendants who were classified as malingering by the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992). However, it is difficult to determine how to interpret these results, as defendants were only administered the SIRS if the evaluating clinician determined that their MMPI-2 results were indicative of symptom feigning. Lewis, Simcox, and Berry (2002) also investigated a sample of criminal defendants using the SIRS as the criterion for defining malingering. However, the authors noted that, to increase the number of feigning participants for comparison, a portion of the selected sample was only administered the MMPI-2 and SIRS "because of clinical suspicion of feigning" (Lewis et al., 2002, p. 172). Utilizing the procedures in the two aforementioned studies makes sense in clinical practice; however, such practice raises the possibility of criterion contamination and could therefore confound accurate prediction models.

Furthermore, studies that investigate the utility of the MMPI-2 validity scales in forensic populations are also rare (Bagby, Rogers, & Buis, 1994; Iverson, Franzen, & Hammond, 1995; Lewis et al., 2002; Wygant et al., 2007). For instance, Wygant et al. (2007) examined the relationship between MMPI-2 validity scales and cognitive symptoms validity measures across civil and criminal forensic settings and found that Infrequency (F) and Infrequency-Psychopathology (F<sub>p</sub>) were related to cognitive response bias in addition to overreported psychopathology in criminal forensic evaluations. Utilizing exploratory factor analysis, Nelson, Sweet, Berry, Bryant, and Granacher (2007) suggested that F and F<sub>p</sub> load on a factor representative of overreporting psychotic/rarely endorsed symptoms among compensation seeking plaintiffs.

Recently, two studies employing known-groups designs (Boccaccini, Murrie, & Duncan, 2006; Toomey, Kucharski & Duncan, 2009) examined archival samples of criminal defendants derived from the same setting who had been administered both the MMPI-2 and SIRS as part of a forensic evaluation. These studies found that the F (Boccaccini et al., 2006; Toomey et al., 2009) and the F<sub>p</sub> (Toomey et al., 2009) scales performed quite well in differentiating between a SIRS-defined malingering group and a nonmalingering group.

To summarize, in a review of the extensive literature of the F and F<sub>p</sub> scales in detecting malingering in both simulation and known-groups designs, Berry and Schipper (2007) noted that while considerable consistency is evident among mean cutting scores, sensitivity, specificity, and effect sizes (Cohen's *ds*) across the two designs, there appears to be a large decline in mean cut scores for F and mean sensitivity for F<sub>p</sub> when moving from simulation to known-groups designs. However, the authors nonetheless noted that, overall, in psychiatric and criminal neuropsychological evaluations, the use of F and F<sub>p</sub> to assess feigning and exaggeration appear well supported (Berry & Schipper, 2007). No study to date, however, has examined the MMPI-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) validity scales using a known-groups design in a criminal forensic setting.

### The MMPI-2 Restructured Form

The MMPI-2-RF is a substantially shortened version of the test designed to represent the clinically significant substance of the

MMPI-2 item pool with a comprehensive set of psychometrically efficient measures. It is linked conceptually and empirically to modern theories and models of psychopathology and personality. The test includes six sets of scales: Validity, Higher-Order, Restructured Clinical (RC), Specific Problem, Interest, and Personality Psychopathology Five scales.

The current investigation focused on the validity of the MMPI-2-RF scales designed to measure symptom overreporting. These indicators include revised versions of three current MMPI-2 scales and a new measure introduced with the MMPI-2-RF. The Infrequent Responses (F-r) scale serves as a general overreporting indicator and comprises 32 items that are rarely endorsed by the test's normative sample (i.e., were answered in the keyed direction by 10% or less of both men and women in the sample).

The Infrequent Psychopathology Responses (F<sub>p-r</sub>) scale is an indicator of overreported symptoms of severe psychopathology. The MMPI-2 F<sub>p</sub> scale, composed of 27 items, was developed originally by Arbisi and Ben-Porath (1995) to complement the F scale, on which scores are confounded by genuine reports of severe psychopathology. F<sub>p-r</sub> is shorter than its MMPI-2 counterpart, consisting of 21 items, 17 of which also are scored on F<sub>p</sub>. Item reduction involved the removal of those items also scored on the MMPI-2 Lie scale, those included on the new Infrequent Somatic Responses (F<sub>s</sub>) scale, and several with wording judged to be ambiguous. Four items were also added to F<sub>p-r</sub> based on multiple regression analyses that indicated that they could contribute incrementally to the scale in the detection of overreported test protocols (Tellegen & Ben-Porath, 2008).

The F<sub>s</sub> scale was added to the MMPI-2-RF to measure overreported somatic complaints using the traditional infrequency approach. Wygant, Ben-Porath, and Arbisi (2004) developed F<sub>s</sub> by identifying 16 items with somatic content that were endorsed by less than 25% of patients in two large archival medical samples and an archival chronic pain sample, comprising over 55,000 patients. Wygant (2007) examined the scale in several simulation, known-groups, and mental health samples and found that it was significantly elevated among patients who failed cognitive symptom validity tests and participants instructed to feign symptoms consistent with a head injury. F<sub>s</sub> was also less correlated with measures of genuine somatic complaints and mood psychopathology than other MMPI-2 validity scales, and the scale added incrementally to other MMPI-2 validity scales in predicting various response bias criteria.

Finally, a revised version of the Symptom Validity (FBS-r) scale assesses noncredible somatic and neurocognitive complaints. Research has found that the original Symptom Validity scale (Lees-Haley, English, & Glenn, 1991),<sup>1</sup> which was developed specifically as a validity scale for use in civil forensic assessments, is sensitive to noncredible somatic responding (e.g., Bianchini, Etherton, Greve, Heinly, & Meyers, 2008; Larrabee, 1998, 2003), suboptimal effort on cognitive symptom validity tests (e.g., Greiffenstein, Baker, Gola, Donders, & Miller, 2002; Larrabee, 2003; Ross, Millis, Krukowski, Putnam, & Adams, 2004; Wygant et al.,

<sup>1</sup> The Symptom Validity Scale (FBS) was formerly labeled the Fake Bad Scale but was renamed when officially adopted on the MMPI-2 scoring materials to better reflect its construct validity and to avoid the negative connotation associated with such a label.

2007), and malingered neurocognitive dysfunction (Greve, Bianchini, Love, Brennan, & Heiny, 2006). The FBS-r comprises 30 of the original 43 FBS items. Whereas the three infrequency scales, F-r, F<sub>P-r</sub>, and F<sub>S</sub>, do not overlap in content, FBS-r shares three items with F<sub>S</sub> and one with F<sub>P-r</sub>.

Ben-Porath and Tellegen (2008) reported correlations between the MMPI-2-RF and MMPI-2 versions of F-r, F<sub>P-r</sub>, and FBS-r showing strong covariation in scores on the original and revised scales (correlation coefficients greater than .90). However, they noted the need to examine the utility and efficacy of various cutoffs on the revised scales in detecting noncredible symptom reporting. To date, only one published study has examined the validity and clinical utility of the MMPI-2-RF validity scales designed to detect overreporting. Wygant et al. (2009) examined these scales in two simulation samples and one known-groups sample that utilized cognitive symptom validity tests as a criterion and found that the scales were able to detect the various threats to protocol validity in civil forensic settings.

### The Current Study

The current investigation was designed to examine the validity and clinical utility of the MMPI-2-RF validity scales in detecting malingered psychopathology in a known-groups design. We used the SIRS (Rogers et al., 1992), a well-validated measure of malingering often considered a gold standard, as the external criterion. Because the SIRS primarily reflects malingering of extreme psychopathology, we expected that the F-r and F<sub>P-r</sub> scales would exhibit greater utility in differentiating between the malingering and nonmalingering groups than would F<sub>S</sub> and FBS-r, which are specifically focused on detecting noncredible somatic and neurocognitive responding. Because of their nonoverlapping nature, somewhat differing detection strategies (rare symptoms [F<sub>P-r</sub>] vs. quasi-rare symptoms [F-r]; see Rogers, 2008), and the fact that F-r is more saturated with emotional content whereas F<sub>P-r</sub> is more associated with bizarre content (Tellegen & Ben-Porath, 2008), we expected that these scales would add incremental predictive utility to one another in detecting malingering. Finally, we also examined the optimal cut scores for these scales in detecting malingered psychopathology and compared them to those proposed in the MMPI-2-RF manual (Ben-Porath & Tellegen, 2008).

### Method

#### Participants

Potential participants consisted of 155 men who were selected from an archival database of criminal defendants referred by the federal courts for competency to stand trial, criminal responsibility, or aid-in-sentencing evaluations between 1994 and 2004.<sup>2</sup> A large portion of these data has been used previously in the examination of MMPI-2 validity scales (Toomey et al., 2009). Each of these individuals had been administered the MMPI-2 and the SIRS (Rogers et al., 1992) as part of their psychological evaluation. Individuals who produced an excessive amount of unscorable responding (i.e., Cannot Say > 17), inconsistent (i.e., Variable Response Inconsistency [VRIN-r] > 79T), or indiscriminant fixed (i.e., True Response Inconsistency [TRIN-r] > 79T) responding

based on MMPI-2-RF scales were excluded from the study. These criteria led to the exclusion of 30 (or 19.4%) individuals.

The final sample consisted of 125 men with a mean age of 36.7 years ( $SD = 10.2$ ). They had an average education of 10.2 years ( $SD = 2.9$ ). In terms of race/ethnicity, participants were Caucasian (51.8%), African American (44.0%), or Hispanic (3.5%), with the remaining participants of other or mixed ethnicities. Almost 64% of the sample reported prior mental health treatment, while 38% reported a history of at least one prior psychiatric hospitalization ( $M = 2.7$ ,  $SD = 7.2$ ). Approximately 84% of participants reported substance abuse problems, and 94% had at least one prior felony conviction. There were no statistically significant differences between those participants included versus excluded based on MMPI-2-RF criteria on any of these descriptive variables.

### Measures

**MMPI-2-RF.** The MMPI-2-RF (Ben-Porath & Tellegen, 2008) is a 338-item self-report inventory on which participants respond "true" or "false" to a variety of statements indicating personality and psychopathology. The entire MMPI-2-RF item pool can be derived from the original MMPI-2 (administered in this sample), and the same normative sample is used with a few modifications (Ben-Porath & Tellegen, 2008). The MMPI-2-RF technical manual provides extensive reliability and validity data for this instrument (Tellegen & Ben-Porath, 2008). In addition, Tellegen and Ben-Porath (2008) presented data indicating that the MMPI-2-RF scale scores derived from administration of the 567-item MMPI-2 booklet are interchangeable with results obtained from administration of the 338-item MMPI-2-RF booklet, which included virtually identical mean scale elevations and correlations with external criteria. For this study, we used the F-r, F<sub>P-r</sub>, F<sub>S</sub>, and FBS-r validity scales. See the earlier description of their development and initial validation. We also used the RC scales for the differential symptoms presentation analyses described later.

**SIRS.** The SIRS (Rogers et al., 1992) was used as the external criterion from which malingering and nonmalingering groups were derived. It is a 172-item structured interview designed to assess malingering and overreported response styles. It contains eight primary scales that are designed for specific detection strategies. These include Rare Symptoms, Improbable and Absurd Symptoms, Symptom Combinations, Blatant Symptoms, Subtle Symptoms, Symptom Severity, Symptom Selectivity, and Reported Versus Observed Symptoms scales. Four classifications can be derived from each scale: honest, indeterminate, probable faking, and definite faking. Several reviews (e.g., Rogers, 2001) have indicated excellent psychometric properties for the SIRS, and it has been frequently used as an external criterion measure in known-groups designs for other structured interviews (e.g., Vitacco, Rogers, Gabel, & Munizza, 2007) and self-report inventories (e.g., Edens, Poythress, & Watkins-Clay, 2007) designed to detect malingering. Because we used an archival database in which we did not have

<sup>2</sup> Previous research from this archival database has indicated that about 83% of all individuals evaluated were administered the SIRS. There was no discernible pattern as to why this instrument was administered versus not administered. There were no significant differences between those administered versus not administered the SIRS on any demographic or mental health variables (see e.g., Boccaccini et al., 2006; Toomey et al., 2009).

access to individual SIRS items, we were unable to calculate reliability estimates for the current sample. However, numerous previous studies have shown acceptable to excellent internal consistencies (range: .77–.96) and excellent interrater reliability (range: .97–1.00) for SIRS scales (e.g., Rogers et al., 1992; Ustad, 1998; Vitacco et al., 2007).

## Procedures

We employed a known-groups design (see Rogers, 2008) to examine the utility of the MMPI-2-RF validity scales in detecting malingering. The SIRS was used as an external criterion to determine group membership (i.e., malingering or nonmalingering). We used a strict cutoff of three or more subscales in the probable range or one score in the definite range to determine malingering, as this cutoff has been well validated in forensic populations (e.g., Rogers, Hinds, & Sewell, 1996). This resulted in 27 participants (or 21.6%) being classified as malingering. The remaining 98 participants were classified as nonmalingering.<sup>3</sup> Moreover, the SIRS manual classifies individuals who score in the probable range on one or two subscales as indeterminate. Eight (or 6.4%) participants in the present sample would fall in this category. Because this group would be too small to be analyzed separately but their presence in the nonmalingering group might violate the assumption of a bona fide known-groups design, we also compared the malingering group to a nonmalingering group ( $n = 90$ ) from which these 8 indeterminate participants were excluded.

The participants represented an archival sample of convenience and were not randomly selected. Each participant had undergone a psychological evaluation for the federal court. All treatment and psychiatric hospitalization information was gathered through pre-trial service reports, prior hospitalization records, and clinical interviews. All psychological testing was conducted by the forensic evaluator taking the ability and cooperation of the defendant into consideration. Those who were not fluent in English, refused specific psychological tests, or produced protocols suggestive of inconsistent responding were excluded as stated earlier. The use of human participants for this study was approved by the Federal Bureau of Prisons and university institutional review boards.

## Results

For all analyses, we used untruncated T scores for MMPI-2-RF scales. Tellegen and Ben-Porath (2008) recommend the use of these scores as artificially truncating validity scale scores (20T at the lower end and 120T at the higher end) and substantive scale scores (20T at the lower end and 100T at the higher end) artificially restricts the range of these variables and, thus, attenuates variance.

### Differential Symptom Presentation

Figure 1 shows the mean profiles for the research conditions (i.e., malingering vs. two nonmalingering groups) on the MMPI-2-RF RC scales. We conducted this analysis with the expectation that malingering participants actually exaggerate their symptom presentation relative to the nonmalingering participants. A Bonferroni-corrected alpha set at  $p < .006$  (i.e., .05/9 comparisons) was applied for each individual comparison. Overall multi-

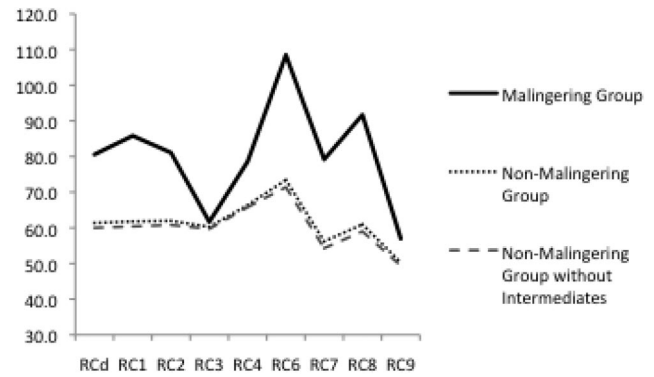


Figure 1. Mean Minnesota Multiphasic Personality Inventory—2 Restructured Form Restructured Clinical (RC) scale profiles for malingering and two nonmalingering groups. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation.

variate analysis of variance (MANOVA) models were significant, Hotelling's  $T = 1.068$ ,  $F(9, 115) = 13.65$ ,  $p < .001$ ,  $\eta^2 = 0.52$ , for the malingering versus nonmalingering group (including indeterminates), and Hotelling's  $T = 1.426$ ,  $F(9, 107) = 16.95$ ,  $p < .001$ ,  $\eta^2 = 0.59$ , for the malingering versus nonmalingering group (excluding indeterminates). The malingering group had significantly higher scores on all of the RC scales (except RC3 [Cynicism] and RC9 [Hypomanic Activation]) relative to both nonmalingering groups, all  $F_s(1, 123) > 18.72$ , all  $p_s < .001$ , all  $\eta^2_s > 0.13$  (vs. nonmalingering group including indeterminates), and all  $F_s(1, 115) > 19.97$ , all  $p_s < .001$ , all  $\eta^2_s > 0.15$  (vs. nonmalingering group excluding indeterminates).

### Validity Scale Group Differences

We next conducted a one-way MANOVA to determine whether there were overall differences on the MMPI-2-RF overreporting validity scales across the two conditions. The overall models were statistically significant—Hotelling's  $T = .983$ ,  $F(4, 120) = 29.49$ ,  $p < .001$ ,  $\eta^2 = .50$  (malingering vs. nonmalingering with intermediates), and Hotelling's  $T = 1.356$ ,  $F(4, 112) = 37.98$ ,  $p < .001$ ,  $\eta^2 = .58$  (malingering vs. nonmalingering without intermediates). We followed up with univariate analyses of variance and Tukey's honestly significant difference post hoc tests for each individual validity scale (i.e., F-r, F<sub>p-r</sub>, F<sub>s</sub>, and FBS-r) to test for differences between the malingering and nonmalingering groups. These results are displayed in Table 1. All four validity scale scores in the malingering condition were significantly higher compared to those of the nonmalingering condition. F-r and F<sub>p-r</sub> had the largest effect sizes, as expected, given that these scales were designed to detect overreported psychopathology. FBS-r had a larger effect size relative to F<sub>s</sub>, which is not surprising given the

<sup>3</sup> We realize that no individual method of detecting malingering is perfect; as such, these groups would be best referred to as *probable malingering* and *probable nonmalingering*. However, for simplicity of writing, we do not use *probable* when referring to these conditions.

Table 1  
Means, Standard Deviations, *F* Tests, and Cohen's *d* Effect Size Estimates for Group Differences

Scale	Nonmalingering groups						<i>F</i> test		Effect size	
	Malingering group ( <i>n</i> = 25)		With intermediates ( <i>n</i> = 98)		Without intermediates ( <i>n</i> = 90)					
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i> <sub>1</sub>	<i>F</i> <sub>2</sub>	<i>d</i> <sub>1</sub>	<i>d</i> <sub>2</sub>
F-r	141.92	23.42	82.00	29.54	79.10	27.37	94.57***	116.55***	2.11	2.37
F <sub>p-r</sub>	122.38	35.54	68.94	22.44	66.37	19.32	91.04***	113.41***	2.07	2.34
F <sub>s</sub>	98.94	25.87	69.17	24.80	67.06	24.02	29.95***	35.29***	1.19	1.30
FBS-r	86.47	14.03	60.97	16.49	59.47	15.88	53.63***	62.98***	1.59	1.74

Note. F-r = Infrequent Responses; F<sub>p-r</sub> = Infrequent Psychopathology Responses; F<sub>s</sub> = Infrequent Somatic Complaints; FBS-r = Symptom Validity; *F*<sub>1</sub> = *F* test between malingering group and nonmalingering group including intermediates; *F*<sub>2</sub> = *F* test between malingering group and nonmalingering group excluding intermediates; *d*<sub>1</sub> = effect size for difference between malingering group and nonmalingering group including intermediates; *d*<sub>2</sub> = effect size for difference between malingering group and nonmalingering group excluding intermediates.

\*\*\* *p* < .001.

broader nature of the overreported symptoms captured by the FBS-r scale. The pattern of results was identical, albeit with slightly larger effect sizes, when intermediate malingering participants were removed from the nonmalingering group.

### Incremental Validity

We next conducted hierarchical logistic regression analyses to determine whether other scales added to F-r or F<sub>p-r</sub> in differentiating between malingering and nonmalingering conditions. We chose these two scales because they were developed to detect overreporting of psychopathology and are associated with the largest effect sizes in this sample. We conducted six separate planned regression analyses. First, we examined whether any of the scales would add incremental predictive utility to F-r. In three regressions, F-r was entered in the first step, and F<sub>p-r</sub>, F<sub>s</sub>, or FBS-r was entered in the second step. Next, three additional analyses were conducted in which F<sub>p-r</sub> was entered in the first step and the remaining three scales in respective second steps. These analyses were repeated when intermediate malingering participants were removed from the nonmalingering groups. Table 2 shows the results of these analyses, which indicate that only F<sub>p-r</sub> added incrementally to F-r in differentiating the malingering and nonmalingering conditions. In the analyses where F<sub>p-r</sub> was entered first, both F-r and FBS-r added significant incremental utility to F<sub>p-r</sub>. This pattern was identical when intermediate malingerers were removed.

### Classification Accuracy

Classification accuracies of the F-r and F<sub>p-r</sub> validity scales in differentiating the two overreporting groups from the patient group were examined using receiver operating characteristic (ROC) curves, which are calculated based on a function of sensitivity and 1 - specificity.<sup>4</sup> Overall predictive performance was assessed using the area under the ROC curve (AUC). In addition, positive and negative predictive powers (PPP and NPP) are indices of diagnostic efficiency in that they provide a probability that the individual is (or is not) engaging in invalid responding given a certain cutoff value. Although predictive powers are more directly

meaningful to clinicians (because they only have access to a test score) and therefore emphasized here, they are heavily influenced by base rates. We therefore provide estimates of predictive powers across three different base rates, (a) base rate of .15, because it represents the lower end of empirically identified overreporting rates (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998; Rogers et al., 1994); (b) base rate of .30, which represents the higher end of empirically derived overreporting rates (e.g., Mittenberg, Patton, Canyock, & Condit, 2002); and (c) base rate of .50, which yields the fewest classification errors (i.e., sensitivity and specificity are maximized), but also because Ardolf, Denney, and Houston (2007) found that negative response bias could occur in over 50% in individuals referred for consecutive criminal evaluations.

We selected a range of possible optimal cutoff scores with an emphasis on reducing false-positive predictions. In most clinical situations, there is an emphasis on minimizing false-positive prediction errors (i.e., 1 - PPP), as clinicians prefer to avoid designating test takers as malingerers unless substantial evidence for such responding is present. This is favorable to avoiding false-negative predictions, as such errors have less potential negative impact on the individual. Bagby et al. (1994) recommended that PPP values should never be lower than .80 (i.e., false-positive prediction error rate of 20%); however, in high-stakes settings such as these, we would argue for even greater PPP values of .90 or higher (cf. Rogers, 2008).

For ROC analyses with F-r, the AUC was .91 (*SE* = .029) for differentiating the malingering and nonmalingering conditions. The AUC increased to .93 (*SE* = .027) when intermediate participants were excluded. The ROC curve of F-r (and relative to F<sub>p-r</sub>) can be seen in Figures 2A and 2B. The upper half of Table 3 displays the classification accuracy statistics, with statistics to the right of each slash mark indicating values after intermediate participants were removed. As seen in Table 3, sensitivity for F-r is

<sup>4</sup> Sensitivity refers to the proportion of malingering participants correctly classified at a particular cut score, whereas specificity refers to the proportion of nonmalingering participants correctly classified below a particular cut score.

Table 2  
Hierarchical Logistic Regression Analysis Results for Predicting Malingering Status

Step	Parameters					$R^2$	$\Delta R^2$
	Model fit $\chi^2(df)$	$w'$	$\chi^2$ change ( $df$ )	$w'$ change			
Malingering group vs. nonmalingering group with intermediates							
F-r entered first							
1. F-r	62.11***	.70				.60	
2a. F <sub>p-r</sub>	65.78***	.73	3.67*	.17		.63	.03
2b. F <sub>s</sub>	62.47***	.71	0.36	.05		.60	.00
2c. FBS-r	62.67***	.71	0.56	.07		.61	.01
F <sub>p-r</sub> entered first							
1. F <sub>p-r</sub>	52.08***	.65				.53	
2a. F-r	65.78***	.73	13.70***	.33		.63	.10
2b. F <sub>s</sub>	53.48***	.65	1.40	.11		.54	.01
2c. FBS-r	59.31***	.69	7.24**	.24		.58	.04
Malingering group vs. nonmalingering group without intermediates							
F-r entered first							
1. F-r	69.50**	.77				.68	
2a. F <sub>p-r</sub>	75.44**	.80	5.94*	.23		.72	.04
2b. F <sub>s</sub>	69.91**	.77	0.42	.06		.68	.00
2c. FBS-r	69.57**	.77	0.07	.02		.68	.00
F <sub>p-r</sub> entered first							
1. F <sub>p-r</sub>	59.67***	.71				.61	
2a. F-r	75.44***	.80	15.77**	.37		.72	.11
2b. F <sub>s</sub>	61.79***	.73	2.12	.13		.62	.01
2c. FBS-r	67.76***	.76	8.10*	.26		.66	.05

Note. Nagelkerke  $R^2$  estimation was used for logistic regression. F-r = Infrequent Responses; F<sub>p-r</sub> = Infrequent Psychopathology Responses; F<sub>s</sub> = Infrequent Somatic Complaints; FBS-r = Symptom Validity;  $w'$  = effect size for  $\chi^2$  statistic;  $w'$  change = effect size for  $\chi^2$  change statistic.  
\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

particularly good, in that it ranges from .89 (at 120T) to .96 (100T), and corresponding NPPs are even higher, indicating that most malingering participants will be identified at these scores. However, as stated earlier, the most important statistic here are PPP values, which, for F-r, do not become acceptable until the base rate reaches .50, with a cut score of 120T providing the best balance of predictive powers. The MMPI-2-RF manual (Ben-Porath & Tellegen, 2008) recommends that overreporting be suspected at F-r  $\geq$  100T, which is supported by the present results in that sensitivity and NPP (across base rates) rates are very high, but so are false-positive predictive error rates as well—an indication that a significant proportion of nonmalingerers score high on F-r.

For ROC analyses with F<sub>p-r</sub>, the AUC was .89 ( $SE = .037$ ; and .91,  $SE = .035$ , when intermediates were excluded) for differentiating the malingering and nonmalingering conditions. Figures 2A and 2B display the ROC curve for F<sub>p-r</sub> relative to F-r. The lower half of Table 3 displays the classification accuracy statistics. As seen in this table, sensitivity and NPP values are lower for F<sub>p-r</sub> compared to F-r, indicating that this scale is less sensitive in correctly identifying malingerers. However, with our stated emphasis on avoiding false-positive prediction errors, the results indicate that a cut score of 110T provides for the best PPP values (and better relative to F-r), but these do not become acceptable until higher base rates. Moreover, when more clean known groups are differentiated (i.e., intermediate participants removed), PPP values for F<sub>p-r</sub> are substantially better, especially at low base rates. Furthermore, the MMPI-2-RF manual recommends that individu-

als who produce 80T or greater on F<sub>p-r</sub> be considered as possibly overreporting (Ben-Porath & Tellegen, 2008). The present results support this recommendation, as such cut scores are associated with high sensitivity and NPP values, but in light of 78% specificity (and even lower PPP values), it will also yield many false-positive prediction errors (up to 60%) in settings where the base rate for overreporting is low.

### Discussion

The goal of this investigation was to determine the utility of the MMPI-2-RF validity scales in detecting malingering in a known-groups design. We found that, as expected, F-r and F<sub>p-r</sub> were the best scales in differentiating malingering and nonmalingering groups as determined by the SIRS. This differentiation was also associated with very large effect sizes, even by malingering research standards ( $d > 1.75$ ; see Rogers, 2008). Moreover, the effect sizes derived from the current study are quite similar to those reported in Rogers et al.'s (2003) meta-analysis of the MMPI-2 validity scales (i.e., mean  $ds = 2.21$  and 1.90 for F and F<sub>p</sub>, respectively), which suggests that F-r and F<sub>p-r</sub> are as effective in differentiating between malingered and nonmalingered profiles as their MMPI-2 counterparts.

The other two MMPI-2-RF validity scales designed to assess overreported response bias, F<sub>s</sub> and FBS-r, performed less well. This finding was entirely expected, however, as F-r and F<sub>p-r</sub> are designed to measure malingering of psychopathology, whereas F<sub>s</sub>

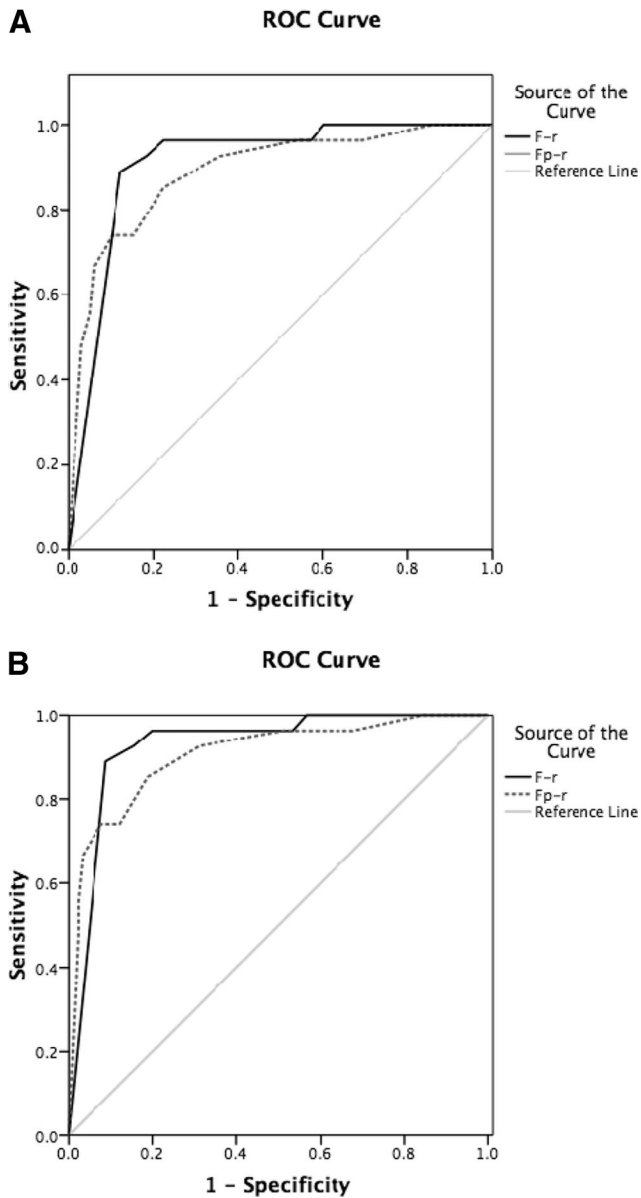


Figure 2. A: Receiver operating characteristic (ROC) curves for differentiating between malingering and nonmalingering (including intermediates) groups. B: ROC curves for differentiating between malingering and nonmalingering (excluding intermediates) groups. Diagonal segment is indicative of chance prediction. F-r = Infrequent Responses; F<sub>p-r</sub> = Infrequent Psychopathology Responses.

and FBS-r are more specifically focused on the detection of noncredible somatic and/or neurocognitive complaints (see Wygant et al., 2009). Nonetheless, F<sub>s</sub> and FBS-r were still associated with respectable effect sizes in differentiating the malingering and nonmalingering groups, suggesting some utility in this setting. This finding is also consistent with the previous observation that individuals in criminal forensic contexts tend to malingering symptoms across all three domains of response bias (i.e., psychopathology, cognitive, somatic), whereas individuals in civil forensic settings tend to be more specific in their symptom presentation (Wygant et al., 2007).

The current findings also indicate that F-r and F<sub>p-r</sub> complement each other in detecting malingering. The regression analyses indicate that both add incremental predictive utility to one another, showing that both scales add something unique in differentiating between malingering and nonmalingering individuals. One potential reason for this finding is that F-r and F<sub>p-r</sub> reflect different response bias detecting strategies (Rogers, 2008). F<sub>p-r</sub> adheres to the more stringent *rare symptoms* strategy, which focuses on identifying symptoms that are rarely reported among genuine patients with mental illness (Rogers, 2008). F-r, on the other hand, corresponds best to the *quasi-rare symptoms* strategy, which focuses on identifying symptoms that are rare in a normal population, and it is *quasi* because individuals with actual psychopathology could endorse these items as well (Rogers, 2008). Thus, these strategies work in a complementary fashion in criminal forensic settings, likely because criminal forensic clients tend to be less sophisticated regarding the symptoms they report, as previously discussed (Wygant et al., 2007).

Although F-r performed quite well in this study, it is important to note a significant caution against solely relying on the quasi-rare symptom detection strategy. Rogers (2008; Rogers et al., 2003) noted that because these symptoms are not necessarily rare in clinical populations, items that are infrequent in normative (rather than clinical) populations may have different endorsement frequency in various clinical groups. This is exemplified by the original F scale, which demonstrated widely varying mean scores in different diagnostic groups (e.g., Rogers et al., 2003). This has significant implication for, among other things, determining optimal cut scores, which have indeed varied greatly for the original F scale (Rogers et al., 2003). However, as indicated earlier, F-r is different from the original F scale in an important way, as it is composed of infrequent items in the current MMPI-2 normative sample, whereas many items on the original F scales were no longer infrequently endorsed in the restandardization sample (Tellegen & Ben-Porath, 2008). Nonetheless, it is imperative that future research continue to examine F-r in various diagnostic groups to determine whether the same variations in mean scores persist. Such research would also have significant implications on deriving stable cut scores to classify malingering, which are discussed next.

Another important goal of this investigation was to determine optimal cut scores on the MMPI-2-RF validity scales to classify malingering versus nonmalingering defendants. We focused on F-r and F<sub>p-r</sub> because these scales were associated with the greatest differentiation in this sample as well as being designed to measure overreporting of psychopathology. When determining optimal cut scores, it is important to consider whether the measures should be used to screen in or screen out malingering. The former would emphasize sensitivity and reducing false-negative prediction errors (i.e., 1 - NPP), but often at a cost of an increased false-positive error rate. Conversely, when the goal is to screen out malingerers, cut scores should emphasize high specificity and low false-positive prediction errors (i.e., 1 - PPP), at the cost of increased non-detection of malingering. Although our goal in recommending cut scores would be to minimize false-positive errors, as we would prefer to avoid designating test takers as malingerers unless substantial evidence for such responding is present, the present data can speak to the scales' utility in both scenarios.



Table 3  
 Classification Accuracy Statistics for F-R and F<sub>p-r</sub> in Differentiating Between Malingering and Nonmalingering Groups

Cutoff score	SN	SP	OCC <sup>a</sup>	BR = .15		BR = .30		BR = .50	
				PPP	NPP	PPP	NPP	PPP	NPP
F-r									
<b>T = 120</b>	<b>.89</b>	<b>.88/.91</b>	<b>.88/.91</b>	<b>.56/.64</b>	<b>.98/.98</b>	<b>.76/.81</b>	<b>.95/.95</b>	<b>.88/.91</b>	<b>.89/.89</b>
T > 115	.93	.82/.84	.84/.86	.47/.51	.98/.98	.68/.72	.96/.96	.83/.86	.92/.92
T > 105	.96	.78/.80	.82/.84	.43/.46	.99/.99	.65/.67	.98/.98	.81/.83	.95/.96
T > 100	.96	.72/.74	.78/.79	.38/.40	.99/.99	.60/.62	.98/.98	.78/.79	.95/.95
F <sub>p-r</sub>									
<b>T &gt; 110</b>	<b>.67</b>	<b>.94/.97</b>	<b>.88/.90</b>	<b>.66/.78</b>	<b>.94/.94</b>	<b>.82/.90</b>	<b>.87/.87</b>	<b>.92/.95</b>	<b>.74/.74</b>
T > 100	.74	.90/.92	.86/.88	.56/.63	.95/.95	.76/.80	.89/.89	.88/.90	.78/.78
T > 90	.74	.85/.88	.81/.85	.46/.52	.95/.95	.67/.72	.88/.89	.83/.86	.77/.77
T > 80	.85	.78/.81	.79/.82	.40/.44	.97/.97	.62/.66	.92/.93	.79/.82	.84/.85

Note. Optimal cut score is set in bold font. Values to the left of a slash are when the nonmalingering group with intermediates is used, whereas values to the right of a slash are when the nonmalingering group without intermediates is used. F-r = Infrequent Responses; F<sub>p-r</sub> = Infrequent Psychopathology Responses; SN = sensitivity; SP = specificity; OCC = overall correct classification; BR = base rate; PPP = positive predictive power; NPP = negative predictive power; T = T score.

<sup>a</sup> OCC values are based on base rates in the current sample (.22 and .23 for nonmalingering groups with and without intermediates, respectively).

The current findings clearly show that the optimal cut score for F-r is 120T in this sample, which is in accordance with the recommended cut score outlined in the MMPI-2-RF manual (Ben-Porath & Tellegen, 2008); however, it is important to keep in mind that even at a high correct classification rate of nonmalingering (88%–91%), when base rates are very low, more than one third of positive predictions will be errors. Thus, clinicians should be careful using F-r as a screening-in tool in low base-rate settings. On the other hand, a cut score of 100T was associated with almost perfect sensitivity, that is, almost all malingerers will score at least 100T on F-r. Thus, this scale could serve as an excellent screening measure of malingering, but clinicians should be aware that this scale has the potential to yield a high number of false-positive errors, especially at these lower cut scores.

The optimal cut score for F<sub>p-r</sub> in this investigation is 110T, which is slightly higher than the manual's recommendation (100T; see Ben-Porath & Tellegen, 2008). F<sub>p-r</sub> was generally associated with less sensitivity than F-r but with higher rates of specificity, indicating that this scale will yield fewer false-positive classifications but also miss more individuals who malingering. Of course, it is important to note that even with high levels of specificity (97% when intermediate participants are removed), PPP values are low at the 15% base rate (.78) by any convention suggested earlier (i.e., .80 or .90), thus further emphasizing the difficulty of using these scales to rule in malingering.

It is important, however, to consider that clinicians would not use these scales in isolation when making decisions about malingering versus nonmalingering. Any test score warrants corroboration by extratest information. As such, although F-r scores of 120T and/or F<sub>p-r</sub> scores of 110T might indicate a substantial likelihood for malingering, other information (e.g., presence of external incentive, no previous mental health treatment history, other test scores) would be needed before malingering could be ruled in. In this light, the classification accuracies of F-r and F<sub>p-r</sub> are impressive.

A final point of discussion is of methodological nature. In the current study, we used a known-groups design with a well-

established external measure (SIRS) as the criterion on which groups were formed. Although the SIRS has excellent specificity and PPP rates, sensitivity rates are lower (Rogers et al., 1992). Several malingering individuals would have been classified as indeterminate by the SIRS and included among nonmalingerers in the current study; therefore, their presence in the nonmalingering group might violate the assumption of a bona fide known-groups design. Consequently, we also analyzed the data without the participants who were classified as indeterminate. The same pattern of findings emerged (albeit somewhat stronger), and it is noteworthy that classification accuracies, especially at low base rates, improved substantially. Future researchers should take these findings into account and weigh the importance of generalizability of findings from a sample to a population that obviously includes intermediates versus need for clean groups with the potential overstatement of scales' utility in detecting malingering.

There are some limitations of this research that need to be acknowledged. First, it is important to note that although the SIRS is currently considered the gold standard for assessing malingering, this external criterion method is not perfect, and it too is associated with some classification error. Although using a cut score of three or more primary scales in the probable feigning or one scale in the definite feigning range is associated with near-perfect specificity, sensitivity has varied in different investigations (Gothard, Viglione, Meloy, & Sherman, 1995; Rogers et al., 1992). This would suggest that our current findings may be an underestimate in detecting malingering, as some actual malingerers may have been in the nonmalingering group. These findings therefore require replication using other research designs (e.g., analogue simulation designs) and other response bias criterion measures.

Moreover, because of the archival nature of the database, we could not calculate reliability estimates for the SIRS, including internal consistency and interrater reliability. However, as indicated earlier, previous research has consistently reported very high reliability estimates, and given the structured response format, these estimates are not surprising, and it is unlikely that significant random measurement error has influenced our results.

In light of these limitations, we also have recommendations for future research. The results of the current investigation warrant replication in samples derived from different contexts (correctional, inpatient, personal injury, etc.). Researchers also need to continue to use other research designs, including analogue simulation and bootstrapping paradigms (see Rogers, 2008). It is also important to examine the utility of these scales in identifying malingering of specific forms of psychopathology with specific diagnostic comparison groups. Such research has particular implications for determining which validity scale works best in different contexts but also for determining the stability of optimal cut scores, particularly with F-r. Finally, the MMPI-2-RF primarily makes use of the rare symptom and quasi-rare symptom detection strategies (cf. Rogers, 2008),<sup>5</sup> which have proved quite successful. Nonetheless, Rogers et al. (2003) showed that the erroneous subtype detection strategy (as approximated with the MMPI-2 Dissimulation scale) is a close second. Therefore, it would be worthwhile to explore whether additional malingering detection strategies would add useful information to the rare symptom strategy in identifying such response bias.

<sup>5</sup> It is unclear to what symptom strategy FBS-r adheres, as it is a derivative of the MMPI-2 FBS, which was developed by examining items that differentiated between malingering and nonmalingering groups in a forensic neuropsychology sample. R. Rogers (personal communication, June 2009) suggested that the FBS adheres to a mixed strategy that is not clearly defined.

## References

- Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale, *F(p)*. *Psychological Assessment*, 7, 424-431.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87, 84-94.
- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clinical Neuropsychologist*, 21, 899-916.
- Bagby, R. M., Rogers, R., & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. *Journal of Personality Assessment*, 62, 191-203.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form; Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Berry, D. T. R., Baer, R. A., & Harris, M. J. (1991). Detection of malingering on the MMPI: A meta-analytic review. *Clinical Psychology Review*, 11, 585-598.
- Berry, D. T. R., & Schipper, L. J. (2007). Detection of feigned psychiatric symptoms during forensic neuropsychological examinations. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 226-263). New York, NY: Oxford University Press.
- Bianchini, K. J., Etherton, J. L., Greve, K. W., Heinly, M. T., & Meyers, J. E. (2008). Classification accuracy of MMPI-2 validity scales in the detection of pain-related malingering: A known-groups study. *Assessment*, 15, 435-449.
- Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal-forensic sample with the Personality Assessment Inventory. *Psychological Assessment*, 18, 415-423.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring* (2nd ed.). Minneapolis, MN: University of Minnesota Press.
- Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, 31, 141-154.
- Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment*, 88, 33-42.
- Gassen, M. D., Pietz, C. A., Spray, B. J., & Denney, R. L. (2007). Accuracy of Megargee's Criminal Offender Infrequency (Fc) Scale in detecting malingering among forensic examinees. *Criminal Justice and Behavior*, 34, 493-504.
- Gothard, S., Viglione, D., Meloy, R., & Sherman, M. (1995). Detection of malingering in competency to stand trial evaluations. *Law and Human Behavior*, 19, 493-505.
- Graham, J. R., Watts, D., & Timbrook, R. E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment*, 51, 264-277.
- Greiffenstein, M. F., Baker, W. J., Gola, T., Donders, J., & Miller, L. (2002). The Fake Bad Scale in atypical and severe closed head injury litigants. *Journal of Clinical Psychology*, 58, 1591-1600.
- Greve, K. W., Bianchini, K. J., Love, J. M., Brennan, A., & Heinly, M. T. (2006). Sensitivity and specificity of MMPI-2 validity scales and indicators to malingered neurocognitive dysfunction in traumatic brain injury. *Clinical Neuropsychologist*, 20, 491-512.
- Iverson, G. L., Franzen, M. D., & Hammond, J. A. (1995). Examination of inmates' ability to malingering on the MMPI-2. *Psychological Assessment*, 7, 118-121.
- Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, 34, 491-498.
- Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *Clinical Neuropsychologist*, 12, 179-188.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *Clinical Neuropsychologist*, 17, 410-425.
- Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, 68, 203-210.
- Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, 14, 170-176.
- Megargee, E. I. (2004, May). *Development and initial validation of an MMPI-2 infrequency scale (Fc) for use with criminal offenders*. Paper presented at the 39th Annual Symposium on Recent Developments on the MMPI-2/MMPI-A, Minneapolis, MN.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102.
- Nelson, N. W., Sweet, J. J., Berry, D. T., R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society*, 13, 440-449.
- Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York, NY: Guilford Press.
- Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York, NY: Guilford Press.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of*

- Reported Symptoms: Professional manual.* Odessa, FL: Psychological Assessment Resources.
- Rogers, R., & Bender, S. (2003). Evaluation of malingering and deception. In A. M. Goldstein & I. B. Weiner (Eds.), *Handbook of psychology: Vol. 11. Forensic psychology* (pp. 109–129). New York, NY: Wiley.
- Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and the SIMS. *Journal of Personality Assessment, 67*, 244–257.
- Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior, 22*, 353–367.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 160–177.
- Rogers, R., Sewell, K. W., & Salekin, R. T. (1994). A meta-analysis of malingering on the MMPI-2. *Assessment, 1*, 227–237.
- Ross, S. R., Millis, S. R., Krukowski, R. A., Putnam, S. H., & Adams, K. M. (2004). Detecting incomplete effort on the MMPI-2: An examination of the Fake-Bad Scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology, 26*, 115–124.
- Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*, 213–223.
- Tellegen, A., & Ben-Porath, Y. S. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Technical manual.* Minneapolis, MN: University of Minnesota Press.
- Toomey, J. A., Kucharski, L. T., & Duncan, S. (2009). The utility of the MMPI-2 Malingering Discriminant Function Index in the detection of malingering: A study of criminal defendants. *Assessment, 16*, 115–121.
- Ustad, K. L. (1998). *Assessment of malingering in a jail referral population: Screening and comprehensive evaluation* (Unpublished doctoral dissertation). University of North Texas, Denton, TX.
- Vitacco, M. J., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand trial patients: A known-groups comparison. *Law and Human Behavior, 31*, 249–260.
- Wygant, D. B. (2007). *Validation of the MMPI-2 Infrequent Somatic Complaints (F<sub>s</sub>) scale* (Unpublished doctoral dissertation). Kent State University, Kent, OH.
- Wygant, D. B., Ben-Porath, Y. S., & Arbisi, P. A. (2004, May). *Development and initial validation of a scale to detect infrequent somatic complaints.* Poster session presented at the 39th Annual Symposium on Recent Developments of the MMPI-2/MMPI-A, Minneapolis, MN.
- Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 Restructured Form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known-group samples. *Archives of Clinical Neuropsychology, 24*, 671–680.
- Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology, 22*, 489–499.

Received June 26, 2009

Revision received October 12, 2009

Accepted October 13, 2009 ■