Examination of the MMPI-2 Restructured Form (MMPI-2-RF) Validity Scales in Civil Forensic Settings: Findings from Simulation and Known Group Samples

Dustin B. Wygant a,*, Yossef S. Ben-Porath b, Paul A. Arbisi c, David T.R. Berry d, David B. Freeman e, Robert L. Heilbronner f

a Department of Psychology, Eastern Kentucky University, Richmond, KY, USA
b Department of Psychology, Kent State University, Kent, OH, USA
c Department of Psychology, Minneapolis VA Medical Center, University of Minnesota, Minneapolis, MN, USA
d Department of Psychology, University of Kentucky, Lexington, KY, USA
e Cal Psych FMT, Los Angeles, CA, USA
f Chicago Neuropsychology Group, Chicago, IL, USA

Accepted 1 September 2009

Abstract

The current study examined the effectiveness of the MMPI-2 Restructured Form (MMPI-2-RF; Ben-Porath and Tellegen, 2008) over-reporting indicators in civil forensic settings. The MMPI-2-RF includes three revised MMPI-2 over-reporting validity scales and a new scale to detect over-reported somatic complaints. Participants dissimulated medical and neuropsychological complaints in two simulation samples, and a known-groups sample used symptom validity tests as a response bias criterion. Results indicated large effect sizes for the MMPI-2-RF validity scales, including a Cohen’s $d$ of .90 for Fs in a head injury simulation sample, 2.31 for FBS-r, 2.01 for F-r, and 1.97 for Fs in a medical simulation sample, and 1.45 for FBS-r and 1.30 for F-r in identifying poor effort on SVTs. Classification results indicated good sensitivity and specificity for the scales across the samples. This study indicates that the MMPI-2-RF over-reporting validity scales are effective at detecting symptom over-reporting in civil forensic settings.

Keywords: MMPI-2-RF; MMPI-2 Restructured Form; Malingering; Forensic evaluation; Medico-legal

Introduction

Civil litigation is often highly conflictual, and the potential for large financial awards presents claimants with an incentive to exaggerate or fabricate symptoms. Likewise, insurance companies and defendants who are potentially liable for providing financial compensation for damages have incentives to identify response bias in plaintiffs as grounds for withholding financial awards. Psychological tests are often used in medico-legal evaluations to provide an objective measure of psychological adjustment (Archer, Stredny, & Zoby, 2006). Tests with embedded validity indicators are particularly useful in such evaluations, where examiners may be challenged to defend their interpretation of psychological test results and to detect non-credible symptom reporting.

The Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) is a 338-item 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. The MMPI-2 Restructured Clinical (RC) scales (Tellegen et al., 2003)
assess the major distinctive core components of the test’s Clinical scales and have been carried over to the MMPI-2-RF. The RC scales are supplemented by broad-band higher-order measures of psychological dysfunction and other scales that focus more specifically on a variety of internalizing, externalizing, and interpersonal characteristics.

The MMPI-2-RF also includes eight validity indicators, two of which are revised versions of the MMPI-2 Variable Response Inconsistency (VRIN-r) and True Response Inconsistency (TRIN-r) scales. The test also contains revised versions of the MMPI-2 Lie scale, now labeled Uncommon Virtues (L-r) and the Correction scale, now labeled Adjustment Validity (K-r), two measures of under-reporting. Sellbom and Bagby (2008) have demonstrated the utility of these under-reporting scales in clinical and non-clinical samples. The MMPI-2-RF also has four over-reporting indicators, which are the focus of the present investigation.

The MMPI-2-RF over-reporting indicators include revised versions of three current MMPI-2 scales and a new measure introduced with the MMPI-2-RF. The Infrquent Responses (F-r) scale serves as a general over-reporting indicator and is comprised 32 items that are rarely endorsed in the MMPI-2-RF normative sample (i.e., were answered in the keyed direction by 10% or less). Unlike the MMPI-2 F scale, which was developed with the original MMPI normative sample, and based on the 1989 normative sample is no longer comprised exclusively of infrequently endorsed items, the F-r scale is more similar to the Fb scale of the MMPI-2, which is composed of items infrequently endorsed in 1989 normative sample.

The Infrequent Psychopathology Responses (Fp-r) scale is an indicator of over-reported symptoms of severe psychopathology. The MMPI-2 Fp 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. The Fp-r scale is shorter than its MMPI-2 counterpart, consisting of 21 items, 17 of which also are scored on Fp. Item reduction involved deletion of items that are also scored on the MMPI-2 L scale, of those slated for inclusion on the new Infrequent Somatic Responses (Fs) scale (described next), and of ones judged to be worded ambiguously. Four items were added to Fp-r based on multiple regression analyses that indicated that they could contribute incrementally to the scale (Tellegen & Ben-Porath, 2008).

The Fs scale was added to the MMPI-2-RF to measure over-reported somatic complaints using the traditional infrequency approach. Wygant, Ben-Porath, and Arbisi (2004) developed Fs 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, totaling over 55,000 patients. Wygant (2007) examined the scale in a several simulation, known-groups, and mental health samples and found that it was significantly elevated among the patients who failed cognitive symptom validity tests (SVTs) and participants instructed to feign head injury symptoms. Fs 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 (Previously labeled the Fake Bad Scale, this measure was re-named Symptom Validity to provide a more descriptive and less inferential label [Ben-Porath, Tellegen, & Graham, 2008].) assesses non-credible somatic and neurocognitive complaints. Research has found that the original FBS scale (Lees-Haley, English, & Glenn, 1991), which was developed specifically as a validity scale for use in civil forensic assessments, is sensitive to exaggerated emotional distress in personal injury settings (e.g., Crawford, Greene, Dupart, Bongar, & Childs, 2006; Lees-Haley et al., 1991), somatic malingering (e.g., Larrabee, 1998, 2003), suboptimal effort on cognitive SVTs (e.g., Greiffenstein, Baker, Gola, Donders, & Miller, 2002; Larrabee, 2003; Ross, Millis, Krukowski, Putnam, &, Adams, 2004; Slick, Hopp, Strauss, & Spellacy, 1996; Wygant et al., 2007), and malingered neurocognitive dysfunction (Greve, Bianchini, Love, Brennan, & Heinly, 2006). The FBS-r is composed of 30 of the original 43 FBS items. Whereas the three infrequency scales, F-r, Fp-r, and Fs, do not overlap in content, FBS-r shares three items with Fs and one with Fp-r.

Tellegen and Ben-Porath (2008) report correlations between the MMPI-2-RF and MMPI-2 versions of F-r, Fp-r, and FBS-r showing strong covariation in scores on the original and revised scales (correlation values in the upper .90). However, they note the need to examine the utility and efficacy of various cutoffs on the revised scales in detecting non-credible symptom reporting. The current investigation provides an opportunity to examine the scales in various medicolegal samples, where examiners need to consider the possibility of over-reporting in general and of neurocognitive and somatic symptoms in particular.

Goal of the Present Study

The present study was designed to examine the utility of the MMPI-2-RF validity scales in detecting various threats to protocol validity in civil forensic settings. Using three different samples, the MMPI-2-RF over-reporting validity scales were examined in both simulation and known-groups designs. The first simulation sample compared a group of patients with a history of head injury with participants instructed to feign symptoms of a head injury within the context of a disability evaluation. The second simulation sample compared medical participants instructed to exaggerate somatic and concurrent emotional complaints with medical patients who had no apparent motivation to over-report their symptoms and completed
testing under standard instructions. The use of head injury and medical patients as control groups in these samples addresses one of the major concerns in simulation research, namely the need to include clinical comparison groups to ensure that participants instructed to simulate symptoms can be distinguished from individuals with genuine pathology (Rogers, 1997).

Complementing the results from the simulation designs, a third sample employed a known-groups design to examine the association of the MMPI-2-RF validity scales with cognitive symptom validity measures. Previous research has suggested that evidence of cognitive response bias in personal injury and neuropsychological settings often co-occur with exaggerated somatic and emotional complaints (Larrabee, 1998, 2003). SVTs, which measure an individual’s effort during neuropsychological testing, were used as a criterion for response bias, given their apparent face validity as measures of memory functioning.

It was hypothesized that scores on the MMPI-2-RF general over-reporting indicator (F-r) would be strongly associated with symptom over-reporting in civil forensic settings, as would be scores on the two validity scales associated with exaggerated somatic and neurocognitive complaints (Fs and FBS-r). Because civil forensic claimants rarely report psychotic psychopathology (Larrabee, 2003; Wygant et al., 2007), it was also hypothesized that effect sizes for Fp-r would be smaller than the other over-reporting validity scales.

Materials and Methods

Participants

Head Injury Simulation. This sample, previously examined by Dearth and colleagues (2005) includes 46 participants. Fliers in waiting rooms of clinical specialists and head injury support group meetings recruited 46 participants with a history of head injury that resulted in a loss of consciousness for > 1 hr. The participants had undergone evaluation at a regional medical center, gave consent to have their medical records reviewed, had no current involvement in compensation seeking, however 60% of these participants had been involved in previous litigation. Moreover, no one had any history of significant substance abuse. Demographic information for the sample can be found in Table 1.

Medical Simulation. This sample, previously presented by Sellbom, Ben-Porath, Graham, Arbisi, and Bagby (2005), is comprised 90 male veteran medical outpatients at the Minneapolis Veterans Affairs Medical Center who had been found to have service-connected physical disabilities but did not have a psychiatric disability and were not receiving psychiatric services. Participation in the study was completely voluntary and was not part of their clinical assessment as patients. The patients presented to the medical center with various medical problems, such as musculoskeletal and vision/auditory problems. Participation was excluded for patients who had previously been involved in a lawsuit or were actively seeking service connection for a medically related injury or condition. Demographic information for the sample can be found in Table 1.

Personal Injury/Disability. This sample, previously examined by Wygant and colleagues (2007), consists of 151 personal injury and disability claimants referred by their insurance company, attorney, or worker’s compensation for a psychological evaluation. The claimants were examined in one of two private practices in either Chicago or Los Angeles. Table 1 provides demographic data. Forty-two percent of the sample claimed emotional disability due to work-related stress, 33% experienced a minor head injury, 16% experienced an orthopedic or musculoskeletal injury, and 7% experienced neurological injuries (other than closed head injuries). Seventy percent of the claimants experienced work-related injuries, whereas 20% were involved in motor-vehicle accidents and the remaining 10% was injured in a variety of other circumstances. Although psychiatric diagnoses were not systematically available for examination, approximately 56% of the sample experienced neurological or physical injuries, whereas 42% experienced emotional problems (mostly mood and anxiety symptoms) as their primary concern.

Table 1. Demographics of the samples

<table>
<thead>
<tr>
<th></th>
<th>Personal Injury/Disability (n = 151)</th>
<th>Head Injury Simulation (n = 46)</th>
<th>Medical Simulation (n = 90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Men</td>
<td>42.4</td>
<td>61</td>
<td>100</td>
</tr>
<tr>
<td>Mean age</td>
<td>43.7 (11.4)</td>
<td>31.7 (12.9)</td>
<td>61.2 (11.4)</td>
</tr>
<tr>
<td>Mean education</td>
<td>13.9 (2.4)</td>
<td>12.5 (2.2)</td>
<td>13.7 (4.6)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Caucasian</td>
<td>65.6</td>
<td>97.8</td>
<td>91.1</td>
</tr>
<tr>
<td>% African American</td>
<td>13.9</td>
<td>2.2</td>
<td>6.7</td>
</tr>
<tr>
<td>% Latino/Latina</td>
<td>15.2</td>
<td>0</td>
<td>1.1</td>
</tr>
</tbody>
</table>
Instruments and Measures

**Minnesota Multiphasic Personality Inventory-2-Restructured Form.** All participants were administered the MMPI-2, however, because all of the items of the MMPI-2-RF are included on the MMPI-2, it is possible to score MMPI-2-RF scales in archival MMPI-2 data sets. Tellegen and Ben-Porath (2008) report results of analyses that establish the equivalence of scale scores produced with the two versions of the instrument.

**Symptom Validity Tests**

**Test of Memory Malingering (TOMM; Tombaugh, 1996)** is a 50-item picture recognition test that has been validated in a number of studies. Rees, Tombaugh, Gansler, and Moczynski (1998) demonstrated that the TOMM had a sensitivity of 1.00 and a specificity of .96 in comparing a group of brain-injured patients instructed to malinger neurocognitive deficits on a neuropsychological evaluation (including the TOMM), and a group of brain-injured patients instructed to put forth their best effort on testing. Research has also found that the TOMM is not significantly impacted by depression (Ashendorf, Constantinou, & McCaffrey, 2004; Rees, Tombaugh, & Boulay, 2001; Yanez, Fremouw, Tennant, Strunk, & Coker, 2006) or laboratory-induced pain (Etherton, Bianchini, Greve, & Ciota, 2005).

**Word Memory Test (WMT; Green, 2003; Green, Allen, & Astner, 2004)** is a commonly used forced-choice verbal recognition measure of effort in forensic neuropsychological settings. Green (2003) found that individuals with severe neurological impairment and injuries performed very well on the WMT (i.e., Mean Immediate Recall = 96.7% correct, SD = 3.5%; Mean Delayed Recall = 95.6% correct, SD = 3.3%). Research indicates that the WMT is very sensitive to effort and insensitive to psychosocial variables (age, gender, race/ethnicity, socio-economic status), intelligence, psychopathology, and neurological impairment (Green et al., 2004).

**Computerized Assessment of Response Bias.** The CARB (Allen, Conder, Green, & Cox, 1997) is an SVT that utilizes a digit recognition paradigm. Green and colleagues (2004) found that performance on the CARB was nearly perfect (M = 98.3% correct, SD = 2.6%) with individuals who have sustained documented brain injuries. Furthermore, in re-evaluating data from Dunn, Shear, Howe, and Ris (2003), Green and colleagues found that the CARB had 84% sensitivity and 100% specificity in differentiating between analogue participants instructed to feign symptoms of a head injury on the CARB and individuals instructed to put forth their best effort.

**Victoria Symptom Validity Test.** The VSVT (Slick, Hopp, Strauss, & Thompson, 1997) is another computer-administered digit recognition test. Slick, Hopp, Strauss, Hunter, and Pinch (1994) examined the VSVT with healthy adult controls, non-compensation-seeking post-concussion patients, and unimpaired participants feigning post-concussion syndrome and found that all control participants (healthy controls and non-litigating patients) performed above cutoffs for malingering (i.e., 100% specificity), whereas 83% of the simulators scored in the questionable or invalid range (i.e., 83% sensitivity).

**Procedures**

Following recommendations in the MMPI-2-RF manual (Ben-Porath & Tellegen, 2008), participants from all samples were removed from analysis if they omitted more than 18 responses to the MMPI-2-RF items or exhibited excessive random (VRIN-r > 80) or fixed (TRIN-r > 80) responding.

Participants in the **Head Injury Simulation** sample were divided into two groups. The malingering group was instructed to feign symptoms of head injury on the MMPI-2 and a battery of neuropsychological tests. They were asked to imagine that they were not able to return to work and were to complete testing to document disability. They were cautioned to avoid obvious dissimulation and detection and were provided with a sheet that detailed coaching methods culled from the Internet. Participants were reimbursed $70 for their participation, with the chance of earning an additional $20 if they fulfilled their instructions without detection.

The malingering group was compared with 23 participants with a history of head injury who completed testing under standard instructions. Participants were reimbursed $70 for their time.

In the **Medical Simulation** sample, participants were asked to initially complete the MMPI-2 under standard instructions and then they were randomly assigned to retake the instrument under either a standard test administration condition or an exaggeration condition and they were reimbursed $20 for completing the MMPI-2 twice. Further, as an added incentive, participants had a 25% chance of earning an additional $25 for following their particular set of instructions. Participants were informed that the resulting MMPI-2 profiles would not be used in any decision regarding their care or disability status. Participants in the exaggeration condition were instructed to exaggerate their physical symptoms and the emotional distress stemming from their physical conditions or medical problems. They were instructed to either pretend that they had problems they did not
have or to exaggerate the problems that they did have without over-doing it to avoid detection. The instructions recommended that they imagine they were applying for a Service Connected disability based on a physical injury incurred while on active duty and that their MMPI-2 results would have an impact on whether the disability rating board believed that they had been injured and whether the injury resulted in problems for them. All participants were administered a post-test questionnaire to determine how well they complied with the scripted instructions.

Participants in the Personal Injury/Disability sample completed the MMPI-2 and cognitive SVT as part of their psychological evaluations. Their scores were retrieved archivally. Participants were divided into two groups, based on their SVT performance. Individuals who passed all the SVTs were assigned to one group (n = 93) and were compared with those who failed any of the SVTs administered during their evaluation (n = 47). After removing invalid protocols, 68 of the participants were administered one SVT, 49 were administered two SVTs, and 23 were administered three SVTs. Over half (55%) of the 47 participants in the failed SVT group exhibited response bias on at least two SVTs. Regarding the individual failure rates on the four SVTs administered, 36% of the 78 participants administered the WMT exhibited response bias on the instrument, 32% of the 79 participants administered the TOMM exhibited response bias, 58% of the 45 participants administered the CARB exhibited response bias, and 21% of the 29 participants administered the VSVT exhibited response bias.

Results

Means and standard deviations of the MMPI-2-RF validity scales for the exaggerated group and head injury controls in the Head Injury Simulation sample are presented in Table 2. Participants in the exaggerated group scored significantly higher than the head injury controls on F-r, \( t(44) = 2.64, p = .011 \), Fp-r, \( t(44) = 2.51, p = .016 \), Fs, \( t(44) = 3.06, p = .004 \); however, the difference between the groups was not significantly different for FBS-r, \( t(44) = 1.42, p = .164 \).

Cumulative scale frequencies for the four validity scales across the exaggerated and head injury control groups for the Head Injury Simulation sample are presented in Table 3. Data from this table can be used to calculate the sensitivity and specificity of the scales at various cut scores. In addition, we also provide the likelihood ratios for the scales at various cutoffs. These ratios can be calculated by dividing the sensitivity by the false alarm rate, which is 1 − specificity. In examining the classification of

### Table 2. Comparison between Head Injury Simulation groups (n = 23) and head injury controls (n = 23) in Head Injury Simulation sample

<table>
<thead>
<tr>
<th></th>
<th>Head Injury Controls</th>
<th>Head Injury Simulation Group</th>
<th>t(44)</th>
<th>p-value</th>
<th>d-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean T-score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>19.9</td>
<td>40.2</td>
<td>2.64</td>
<td>.011</td>
<td>.78</td>
</tr>
<tr>
<td>F-r</td>
<td>66.5</td>
<td>91.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fp-r</td>
<td>54.6</td>
<td>77.3</td>
<td>2.51</td>
<td>.016</td>
<td>.74</td>
</tr>
<tr>
<td>Fs</td>
<td>61.7</td>
<td>90.8</td>
<td>3.06</td>
<td>.004</td>
<td>.90</td>
</tr>
<tr>
<td>FBS-r</td>
<td>54.2</td>
<td>64.6</td>
<td>1.42</td>
<td>.164</td>
<td>.42</td>
</tr>
</tbody>
</table>

Notes: Cohen’s d calculated for effect size. F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity.

### Table 3. Frequencies in the Head Injury Simulation sample

<table>
<thead>
<tr>
<th>T-score</th>
<th>F-r</th>
<th></th>
<th></th>
<th>Fp-r</th>
<th></th>
<th></th>
<th>Fs</th>
<th></th>
<th></th>
<th>FBS-r</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% HIC</td>
<td>% ORG</td>
<td>LR</td>
<td>% HIC</td>
<td>% ORG</td>
<td>LR</td>
<td>% HIC</td>
<td>% ORG</td>
<td>LR</td>
<td>% HIC</td>
<td>% ORG</td>
<td>LR</td>
</tr>
<tr>
<td>120</td>
<td>0</td>
<td>26.1</td>
<td>13.0</td>
<td>0</td>
<td>17.4</td>
<td>6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>4.3</td>
<td>43.5</td>
<td>10.1</td>
<td>17.4</td>
<td>43.5</td>
<td>10.1</td>
<td>0</td>
<td>43.5</td>
<td>10.1</td>
<td>0</td>
<td>43.5</td>
<td>10.1</td>
</tr>
<tr>
<td>100</td>
<td>4.3</td>
<td>43.5</td>
<td>10.1</td>
<td>21.7</td>
<td>4.3</td>
<td>43.5</td>
<td>10.1</td>
<td>0</td>
<td>43.5</td>
<td>10.1</td>
<td>0</td>
<td>43.5</td>
</tr>
<tr>
<td>90</td>
<td>8.7</td>
<td>43.5</td>
<td>5.0</td>
<td>21.7</td>
<td>17.4</td>
<td>56.5</td>
<td>3.2</td>
<td>8.7</td>
<td>26.1</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>26.1</td>
<td>56.5</td>
<td>2.2</td>
<td>0</td>
<td>43.5</td>
<td>21.7</td>
<td>60.9</td>
<td>2.8</td>
<td>17.4</td>
<td>43.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>47.8</td>
<td>60.9</td>
<td>1.3</td>
<td>43.5</td>
<td>10.1</td>
<td>30.4</td>
<td>69.6</td>
<td>2.3</td>
<td>17.4</td>
<td>47.8</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>56.5</td>
<td>73.9</td>
<td>1.3</td>
<td>13.0</td>
<td>60.9</td>
<td>4.7</td>
<td>39.1</td>
<td>69.6</td>
<td>1.8</td>
<td>39.1</td>
<td>60.9</td>
<td>1.6</td>
</tr>
<tr>
<td>50</td>
<td>78.3</td>
<td>73.9</td>
<td>0.9</td>
<td>73.9</td>
<td>3.9</td>
<td>73.9</td>
<td>1.2</td>
<td>60.9</td>
<td>69.6</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>100</td>
<td>1.0</td>
<td>100</td>
<td>100</td>
<td>1.0</td>
<td>100</td>
<td>1.0</td>
<td>65.2</td>
<td>73.9</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>100</td>
<td>1.0</td>
<td>100</td>
<td>100</td>
<td>1.0</td>
<td>65.2</td>
<td>73.9</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Cumulative percentages in descending order. HIC = Head Injury Controls; ORG = Over-Reporting Group; LR = Likelihood ratios; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity.
malingering, these ratios reflect the likelihood that an individual with a positive test result is malingering in comparison to the likelihood that an individual with a positive test result is not malingering (Larrabee, 2008).

Cutoffs for F-r were optimal at T-score values of 90 and 100, which produced a sensitivity of .44, and specificity ranging from .91 to .96, along with a likelihood ratio of 5.0 (90T) and 10.1 (100T). Cutoffs for Fp-r were optimal at 80, which produced a sensitivity of .44 and 1.00 specificity. Cutoffs for Fs were optimal at 90 and 100, which produced sensitivity between .57 and .44, and specificity ranging between .83 and .96, with a likelihood ratio of 3.2 (90T) and 10.1 (100T). Finally, cutoffs for FBS-r were optimal between 80 and 90, which yielded a sensitivity ranging between .44 and .26, with a specificity ranging between .83 and .91, and a likelihood ratio of 2.5 (80T) and 3.0 (90T).

Means and standard deviations for the Medical Simulation sample are presented in Table 4. Participants in the over-reported group scored significantly higher than the medical controls on F-r, \(t(73) = 8.86, p < .001\), Fp-r, \(t(73) = 7.74, p < .001\), Fs, \(t(73) = 8.47, p < .001\), and FBS-r, \(t(73) = 9.91, p < .001\).

Cumulative scale frequencies for the Medical Simulation sample are presented in Table 5. Optimal T-score cutoffs for F-r ranged between 80 and 100, which produced sensitivity between .75 and .63, and specificity ranging between .91 and 1.00, with a likelihood ratio ranging between 8.2 (80T) and 14.6 (90T). A cutoff of 70 was optimal for Fp-r, and produced a sensitivity of .72 with a specificity of .95 and a likelihood ratio of 16.0. Cutoffs for Fs were optimal between 80 and 100, which produced...
sensitivity between .75 and .56, with a specificity that ranged between .93 and 1.00, and likelihood ratios of 11.0 (80T) and 29.9 (90T). Cutoffs for FBS-r were optimal between 70 and 90, which produced a sensitivity that ranged between .88 and .41, with a specificity that ranged between .91 and 1.00, and a likelihood ratio of 9.6 (70T) and 14.6 (80T).

Finally, means and standard deviations for the Personal Injury/Disability sample are presented in Table 6. An ANOVA was calculated between those participants who passed all of their cognitive SVTs, those who failed one SVT, and those who failed two to three SVTs. As expected, all four scales increased significantly across the three SVT performance levels. Effect sizes were calculated between those participants who passed their SVTs and those who failed two to three SVTs. Cohen’s $d$ values were highest for F-r ($d = 1.60$), followed by FBS-r ($d = 1.42$), Fs ($d = 1.38$), and Fp-r ($d = 1.21$).

Given that some participants in the failed symptom validity group exhibited poor performance on multiple indicators, we calculated the correlation between the four MMPI-2-RF validity scales and the number of SVTs failed (0–3). F-r was correlated with number of SVTs failed at .53, $p < .001$, Fp-r at .39, $p < .001$, Fs at .49, $p < .001$, and FBS-r at .50, $p < .001$.

Cumulative scale frequencies for the Personal Injury/Disability sample are presented in Table 7. To represent a clearer probable malingering group, we only presented the scale frequencies of those who passed their SVTs and those who failed two to three SVTs. Larrabee (2008) indicated that failure of two SVTs provides a much higher likelihood of malingering rather than relying on the failure of a single SVT. T-score cutoffs for F-r in detecting failure of two to three SVTs were optimal between 90 and 110, which produced sensitivity between .39 and .31 in detecting SVT failure, with a likelihood ratio of 5.1 at 90T and 10.8 at 100T. These cutoffs yielded specificity between .92 and 1.00 for those who passed all SVTs. Sensitivity for Fp-r was optimal at cutoffs between 60 and 70, ranging between .39 and .23 when two to three SVTs were failed, with likelihood ratios of 4.0 (60T) and 10.5 (70T). These cutoffs exhibited excellent specificity, ranging between .92 and .97 when the participants passed all SVTs. Cutoffs for Fs were optimal between 80 and 90, which produced sensitivity between .62 and .35 when two to three SVTs were failed, with a likelihood ratio of 7.2 (80T) and 6.4 (90T). These cutoffs also yielded excellent specificity, ranging between .91 and .95. Finally, a cutoff of 90 for FBS-r produced the optimal classification, with a sensitivity of .39 when failing two to three SVTs, with a likelihood ratio of 7.1. This cutoff yielded excellent specificity of .95.

**Discussion**

Using simulation and known-groups samples, results from the current study suggest that the MMPI-2-RF validity scales are useful in detecting symptom exaggeration associated with medico-legal settings.

Three of the four validity scales were significantly higher in the group instructed to feign symptoms associated with disability compared with the head injury control group in the Head Injury Simulation sample, although Fs and F-r had the largest effect sizes, which was consistent with our hypothesis that dissimulating participants would over-report somatic and emotional complaints to a larger extent than symptoms of severe psychopathology. As evident from the results in Table 3, a cutscore of 100 on F-r and Fs had the best ratio of sensitivity to exaggerated head injury complaints (.44), with a false-positive rate of only .04.

Similarly in the Medical Simulation sample, all four validity scales were significantly elevated in the malingering group. Consistent with our hypotheses, effect sizes were largest for FBS-r, F-r, and Fs, reflecting exaggerated somatic, neurocognitive, and emotional complaints that would be expected in this type of setting. Scale frequency results from Table 5 indicate that

**Table 7. Frequencies in Personal Injury/Disability sample**

<table>
<thead>
<tr>
<th>T-score</th>
<th>F-r</th>
<th>Fp-r</th>
<th>Fs</th>
<th>FBS-r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Pass</td>
<td>% Fail</td>
<td>LR</td>
<td>% Pass</td>
</tr>
<tr>
<td>120</td>
<td>19.2</td>
<td>30.8</td>
<td>7.7</td>
<td>0</td>
</tr>
<tr>
<td>110</td>
<td>34.6</td>
<td>10.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>17.2</td>
<td>38.5</td>
<td>3.1</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>17.2</td>
<td>38.5</td>
<td>3.1</td>
<td>1</td>
</tr>
<tr>
<td>80</td>
<td>7.5</td>
<td>38.5</td>
<td>5.1</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>17.2</td>
<td>61.5</td>
<td>3.6</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>30.8</td>
<td>7.7</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>30.8</td>
<td>7.7</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>1.0</td>
<td>1.0</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>1.0</td>
<td>1.0</td>
<td>100</td>
</tr>
</tbody>
</table>

**Notes:** Cumulative percentages in descending order. PASS = Passed all SVT ($n = 93$); FAIL = Failed 2–3 SVT ($n = 26$); LR = Likelihood ratios; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity.
cutscores vary between scales. A cutscore between 80 and 90 on FBS-r produces the highest sensitivity (0.61–0.66) for exaggerated complaints with a small possibility of false positives (0.0–0.05), whereas, cutscores between 90 and 100 on F-r and Fs produce the best combination of sensitivity (ranging from 0.63 to 0.66 for F-r, and 0.56 to 0.69 for Fs) and specificity (ranging from 0.95 to 1.0 for F-r and 0.98 to 1.0 for Fs).

Finally, in the Personal Injury/Disability sample, in which participants were classified based on their SVT performance, all four scales were significantly higher in the group that evidenced feigned cognitive impairment. Similarly to the Medical Simulation sample, effect sizes were largest for FBS-r, F-r, and Fs, indicating that individuals feigning cognitive deficits over-report somatic, neurocognitive, and emotional complaints on the MMPI-2-RF. Moreover, the scales exhibited expected association with the number of SVTs failed, with F-r, FBS-r, and Fs showing higher correlations than Fp-r. Scores between 80 and 90 on FBS-r, and 90 and 100 on F-r and Fs should alert the examiner to the possibility of feigned neurocognitive deficit in the context of a neuropsychological or disability examination, with FBS-r exhibiting good sensitivity at that range of scores (0.38–0.77), whereas F-r and Fs provided good specificity, ranging from 0.92 (F-r 90T) to 0.99 (Fs 100T). Consistent with previous research (e.g., Larrabee, 2008), the sensitivity of the over-reporting scales to cognitive response bias increased when the participants failed multiple SVTs.

Results from these analyses indicate that Larrabee (2003) and Wygant and colleagues’ (2007) conclusion regarding the Fp scale extends to the Fp-r scale. Fp-r is less sensitive to response bias in various medico-legal settings, where the demand characteristics of malingering typically do not influence the individual to “act crazy” (Larrabee, 2003, p. 674).

Whereas the present study provided an initial examination of the MMPI-2-RF validity scales in civil forensic situations, further research will be needed in samples where patients present with different forms of response bias. Wygant and colleagues (2007) concluded that the demand characteristics of malingering differ across various forensic domains (i.e., criminal vs. civil). Criminal defendants for instance are likely to feign severe psychopathology in order to present themselves as “insane” or incompetent.

Civil forensic settings, such as the ones examined in this study, present the examiner with several challenges. For instance, detecting somatic response bias is complicated by several differential diagnoses with similar clinical presentations (e.g., somatoform and factitious disorders [DSM-IV-TR; APA, 2000]). Examination of the MMPI-2-RF in samples where patients present with physical concerns and do not have an inherent incentive to distort their self-report will be necessary to determine the extent to which scales such as Fs and FBS-r reflect response bias relative to somatization, which is an unconscious expression of somatic complaints in an effort to manage stress or reduce conflict. Preliminary results from Sellbom, Wygant, and Bagby (2008) suggest that Fs may be particularly well suited in this differentiation.

Given the relatively small samples upon which classification of the scales was examined, it is recommended that cutoffs be cross-validated in larger samples. The MMPI-2-RF validity scales should also be examined in relation to the structured malingering criteria proposed by Slick, Sherman, and Iverson (1999) to measure malingered neurocognitive deficits and Bianchini, Greve, and Glynn (2005) for malingered pain-related disability. Both of these criteria were proposed to increase the robustness of malingering diagnosis by incorporating multiple sources of information and conceptualizing malingering as existing on a dimension, ranging from possible malingering, probable malingering, and definite malingering. Both structured criteria propose that self-report measures with built-in validity indicators provide additional evidence to the patient’s physical exam and test results and cognitive SVTs in support of a diagnosis of malingering. Given the MMPI-2-RF validity scales’ ability to detect response bias across the three broad domains of response bias (psychopathology, cognitive impairment, and physical complaints), they appear particularly well suited to assist in the detection of these structured malingering criteria. Preliminary results by Wygant, Gervais, and Ben-Porath (2007) suggested that malingered neurocognitive dysfunction is associated with elevations on F-r, Fs, and FBS-r in a sample of non-head injury disability claimants.

Several limitations need to be noted regarding the samples in this study. As previously mentioned by Dearth and colleagues (2005), the Head Injury Simulation sample has a limited number of participants and subjects were provided only modest financial incentives for successful malingering compared with real-world settings. Also noteworthy, the head injury sample was not obtained consecutively in a clinical setting, but rather, recruited from self-help groups and 60% of the sample had previously been involved in litigation. This may bias the sample to contain individuals who have experienced continuing problems from their injuries. Further examination of the MMPI-2-RF with individuals who have fully recovered from their injuries is warranted as a clinical comparison. Small sample sizes and limited reimbursement were also problematic in the Medical Simulation sample.

Despite the limitations of these samples, the overall results of this study indicate that the MMPI-2-RF validity scales can detect various threats to protocol validity in civil forensic settings. F-r is effective in the general detection of over-reporting, and particularly of exaggerated emotional complaints. Both Fs and FBS-r contribute to the detection of exaggerated somatic and neurocognitive complaints, while capitalizing on different scale construction strategies. Fs utilizes the infrequency
approach that generally emphasizes specificity, whereas FBS-r was constructed with a combined rational and empirical approach that emphasizes sensitivity.

Whereas the Fp-r scale generally had the lowest effect size, compared with the other MMPI-2-RF over-reporting scales in the present study, this is likely due to the demand characteristics of the civil forensic settings in general. In contrast, in civil or criminal forensic settings where the demand characteristics pull to feign severe psychopathology such as PTSD or psychosis, Fp-r will likely be more effective in detecting symptom over-reporting (Arbisi, Ben-Porath, & McNulty, 2006; Wygant et al., 2007).

**Funding**

This study was funded in part by a grant from the University of Minnesota Press, publisher of the MMPI-2-RF. Other partial funding was provided by the Graduate School of the University of Kentucky.

**Conflict of Interest**

YB-P is a paid consultant to the MMPI-2-RF publisher, the University of Minnesota Press, and distributor, Pearson. He received royalties on sales of MMPI-2-RF materials. PAA also receives research funding from the University of Minnesota Press.

**References**


