The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context

Dustin B. Wygant, Martin Sellbom, Yossef S. Ben-Porath, Kathleen P. Stafford, David B. Freeman, Robert L. Heilbronner

a Kent State University, United States
b Psycho-Diagnostic Clinic, Akron, OH, United States
c Comprehensive Psychological Services, UCLA Medical Center, United States
d Chicago Neuropsychology Group, United States

Abstract

The association between scores on MMPI-2 scales and cognitive symptom validity test (SVT) failure was investigated in 127 criminal defendants evaluated for competency to stand trial, criminal responsibility, and drug dependence, and 141 personal injury and disability claimants. Results indicated that SVT failure was associated with exaggerated symptom presentation involving somatic complaints in civil litigants and more global exaggeration of psychopathology and somatic complaints in criminal defendants. Scores on the MMPI-2 Fake Bad Scale (FBS) were associated with SVT failure in both civil and criminal litigants, whereas scores on the MMPI-2 Fp scale were associated with SVT failure in criminal defendants, but not in civil plaintiffs. These results support the utility of the FBS as an indicator of non-credible presentation of somatic and cognitive complaints in both civil and criminal forensic psychological assessments, and indicate that the lack of association between the MMPI-2 infrequency scales and SVT failure is limited to civil forensic settings.

Keywords: Malingering; MMPI-2; Symptom validity test; Word memory test; Test of memory malingering; Forensic evaluation

Malingering is a psychiatric diagnosis that involves the intentional over-reporting or feigning of symptoms that is externally motivated by secondary gain. The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association, 2000), defines malingering as the conscious exaggeration or fabrication of physical or psychological symptoms for external incentives, such as evading responsibility in a criminal case due to mental illness (e.g., insanity) or financial compensation in a disability claim. Research suggests that malingering typically occurs in three broad domains: psychopathology, cognitive impairment, and physical or medical illness (Rogers & Bender, 2003). The differential manifestation of malingering, or response bias, across these three domains has required the development of a variety of techniques for detecting when it occurs.

Rogers (1997) and Rogers and Bender (2003) have detailed the methods developed to detect the malingering of psychopathology, which are primarily premised on the notion that those feigning mental illness do not have a
sophisticated understanding of the intricacies of psychopathology. Two such strategies employed in self-report measures include the endorsement of uncommon symptoms among patients with genuine dysfunction (i.e., rare symptoms approach) or the endorsement of erroneous stereotypes of psychopathology.

Just as accurate self-report is necessary for valid measurement of psychopathology, adequate effort on the part of the test taker is required to ascertain cognitive impairment in neuropsychological testing (Bianchini, Mathias, & Greve, 2001; Rogers & Bender, 2003). Neuropsychological testing provides data on cognitive functioning by testing the limits of an individual’s ability. To accurately assess these limits, the test taker must put forth his or her best effort. When examinees do not perform at their maximum ability during testing, test results will confound functional deficits and poor effort, resulting in an underestimate of cognitive functioning or an overestimate of cognitive impairment.

For the past two decades, the most common paradigm for detecting feigned cognitive impairment involves forced-choice symptom validity testing (SVT; Pankratz, 1983). The concept behind this paradigm involves the presentation of a task that appears on its face to be measuring complicated cognitive processes. These tests appear to measure memory, a cognitive function sometimes compromised by neurological damage; however, the tests actually require only brief, passive attention, which is often resilient to neurological injuries and conditions (for a review, see Bianchini et al., 2001). Patients who exaggerate claims of neurological problems may intentionally perform below their actual ability level on these measures because they misperceive them as being difficult measures of memory functioning. However, individuals with bona fide cognitive impairment generally perform at nearly perfect levels on these tests (Guilmette, Hart, Guiliano, & Leninger, 1994). Green, Rohling, Lees-Haley, and Allen (2001) demonstrated that measures of effort were significantly correlated with neuropsychological test battery scores and accounted for poor performance on these measures to a greater extent than the severity of brain dysfunction.

Research has consistently shown that patients who fail symptom validity tests often report a preponderance of somatic complaints on the MMPI-2 (Boone & Lu, 1999; Larrabee, 1998; Youngjohn, Burrows, & Erdal, 1995). Larrabee (2003) found that compensation-seeking patients exhibiting malingered neurocognitive difficulties, as evidenced by worse than chance performance on the Portland Digit Recognition Test (PDRT; Binder & Willis, 1991), scored significantly higher than moderate-to-severe closed head injury patients on Scales 3 ($d = 1.74$), 2 ($d = 1.60$), 1 ($d = 1.40$), 7 ($d = 1.39$), and 8 ($d = 1.22$) of the MMPI-2. These scales contain items describing physical malfunctioning, concern with health, and neurological complaints, such as problems with attention, concentration and mental confusion. Wygant, Gervais, and Ben-Porath (2005) examined the Restructured Clinical (RC; Tellegen et al., 2003) scales in a sample of over 700 disability claimants who were administered the Word Memory Test (WMT; Green, 2003; Green, Allen, & Astner, 1996; Green & Astner, 1995) and found that patients demonstrating poor performance on the WMT had their highest RC scale elevation on RC1 (Somatic Complaints).

Larrabee (1998, 2003) observed that the traditional validity scales of the MMPI-2 used to gauge exaggerated responding (e.g., infrequency [F], back infrequency [FB], and infrequency psychopathology [FP]) are less sensitive to poor effort on symptom validity tests and to somatic malingering than to exaggerated reports of psychopathology because their items generally lack somatic content. Research has found that the Fake Bad Scale (FBS; Lees-Haley, English, & Glenn, 1991), which was developed specifically as a validity scale in personal injury settings, is sensitive to suboptimal effort on symptom validity tests (Larrabee, 2003; Ross, Millis, Krukowski, Putnam, & Adams, 2004; Slick, Hopp, Strauss, & Spellacy, 1996) and malingered neurocognitive dysfunction (Greve, Bianchini, Love, Brennan, & Heinly, 2006). Based on these and other studies documenting the utility of the FBS, this measure was recently added to the standard MMPI-2 validity scales.1

1. Forensic contexts

The domains of cognitive and physical impairment malingering have often been examined together within the context of civil forensic settings (e.g., personal injury and disability evaluations), given that the external incentive to malingering in these evaluations typically involves financial compensation for injuries resulting in physical and/or cognitive functional deficits or impairment (e.g., back pain, difficulty with concentration or memory following a motor vehicle accident). In criminal forensic evaluations, however, the external incentive to malingering typically involves the evasion of criminal responsibility or delay in trial proceedings. In competency to stand trial and criminal responsibility

---

1 Ben-Porath and Tellegen (2006) describe the review process and outcome that led to this decision.
(i.e., “sanity”) evaluations, for example, successful malingering of symptoms can result in delaying the onset of a trial or result in hospitalization rather than incarceration. Descriptions of psychotic symptoms are often incorporated into the language that defines competency and criminal responsibility/sanity. For example, statutory language describing criminal responsibility often states that the defendant have some form of “severe mental disease” (e.g., psychosis) as requirements for the defense (Rogers & McKee, 1995). Given that this statutory language implies that the defendant possesses a severe mental illness to meet the threshold for insanity (Melton, Petrila, Poyhress, & Slobogin, 1997), malingering within this context would most likely involve the fabrication or exaggeration of psychotic or other severe symptoms of psychopathology. Indeed, Bagby, Rogers, and Buis (1994) demonstrated that individuals feigning psychopathology within the context of an insanity evaluation had their highest elevations on MMPI-2 scales indicative of severe mental illness (i.e., Scales 6 and 8) and over-reported psychopathology (i.e., F scale), in comparison to both controls and forensic inpatients.

Although the use of symptom validity testing in neuropsychological and civil forensic settings (e.g., personal injury evaluations) is quite common and has been examined extensively, relatively little research has examined these measures in criminal forensic settings. Duncan and Ausborn (2002) found a significant correlation between the Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994), a post-hoc symptom validity measure created with the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997a) and the Wechsler Memory Scales-Third Edition (WMS-III; Wechsler, 1997b), and the MMPI-2 F scale in a sample of criminal forensic pretrial defendants. Their study presented some preliminary data suggesting that measures of feigned cognitive impairment are related to over-reported psychotic symptoms on the MMPI-2. However, the study was limited by criterion contamination in that the MMPI-2 was used to initially establish malingering groups for the examination of the RDS. Finally, Lanyon (2001) examined the factor structure of various measures indexing over-reporting, under-reporting, and somatization, and found similar factor structure in both civil and criminal forensic samples. However, his study did not examine how the different demand characteristics in each setting would have influenced performance on these measures.

2. The current study

If, as just argued, malingering entails a different set of demand characteristics in criminal than civil forensic settings (i.e., the feigning of psychotic, rather than somatic and cognitive complaints), the reduced sensitivity of the MMPI-2 infrequency scales to cognitive symptom validity test indictors of suboptimal effort may be a function of the evaluation setting. The present study provides an empirical test of this hypothesis by examining whether the relation between MMPI-2 scores and symptom validity tests varies as a function of forensic setting (i.e., criminal versus civil). Specifically, we hypothesized: (1) that criminal defendants who fail symptom validity tests would produce marked elevations on MMPI-2 scales that measure psychotic symptoms (i.e., the Restructured Clinical Scale Aberrant Experiences, [RC8]), whereas criminal forensic litigants (evaluated for personal injury or disability) who fail symptom validity testing would produce marked elevations on MMPI-2 scales measuring somatic symptoms. (i.e., the Restructured Clinical Scale Somatic Complaints, [RC1])²; and (2) that criminal defendants who fail symptom validity testing will produce significant elevations on Fp, given the hypothesized relation between over-reported psychotic symptoms and malingering in this setting,³ whereas the civil forensic patients who fail symptom validity testing will produce significant elevations on the Fake Bad Scale, as previous research has found strong correlations between somatic over-reporting and feigned cognitive impairment (e.g., Greiffenstein, Baker, Gola, Donders, & Miller, 2002; Larrabee, 1998, 2003; Ross et al., 2004; Wygant et al., 2005). Although previous research has documented a substantially reduced association between the MMPI-2 infrequency scales and SVT failure in civil settings, the association between SVT failure and somatic complaining in criminal forensic settings is unknown. We, therefore, did not have an a-priori hypothesis about the performance of individuals who fail SVTs in criminal settings on MMPI-2 measures of somatic complaining (i.e., RC1) and non-credible somatic complaining (FBS)

² We focused on MMPI-2 scales that are homogeneous measures of the target constructs (i.e., somatization and psychosis), which are better captured by the RC than the clinical scales.
³ The F scale was not utilized as an indicator of psychotic over-reporting because research has found that it is confounded with genuine psychopathology, especially among patients with severe mental illness (Arbisi & Ben-Porath, 1995). Fp was chosen as the sole indicator of psychotic over-reporting because it was developed to address this concern.
3. Methods

3.1. Participants

3.1.1. Criminal forensic
This sample consists of 127 criminal defendants referred for psychological evaluations to a certified forensic center providing psychological evaluations to five criminal court districts in northeastern Ohio. Table 1 provides descriptive demographic statistics for the sample. Overall, the defendants were predominantly single, Caucasian males with slightly less than a high school education. Thirty-eight percent of the sample was referred for a competency to stand trial evaluation, 31% were referred for both competency to stand trial and sanity evaluations, 25% were referred to determine suitability for a drug treatment diversion program, and 4% were referred for sanity evaluations. Regarding psychiatric diagnosis, 69% met criteria for either a substance abuse or dependence disorder, 44% met criteria for a personality disorder (predominantly Antisocial Personality Disorder), 19% met criteria for a mood or anxiety disorder, 14% met criteria for a psychotic disorder (e.g., Schizophrenia, Schizoaffective Disorder, Delusional Disorder), and 14% were diagnosed with Malingering.

Three MMPI-2 validity scales were used to remove protocols marked by non-responding (Cannot Say ≥ 30), fixed inconsistent (TRIN T ≥ 80), and random responding (VRIN T ≥ 80). Thirteen defendants were removed because of this evidence of non-content based responding. Excluded defendants were significantly younger (M = 28.4, S.D. = 6.6) than included defendants (M = 34.8, S.D. = 10.7), t(124) = 2.11, p = .037. There were no significant differences between those included and excluded in the criminal forensic sample on any remaining demographic variables. With regard to symptom validity testing, 78 of the defendants were administered the Test of Memory Malingering (TOMM; Tombaugh, 1996) and 65 were administered the WMT. Sixteen defendants were administered both SVTs.

3.1.2. Civil forensic
This sample consists of 141 personal injury or 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. Overall, the sample is predominantly female, Caucasian, with approximately 14 years of education. 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, 20% were involved in motor-vehicle accidents and the remaining 10% of the sample was injured in a variety of other circumstances. Although psychiatric diagnoses were not systematically available for examination in the civil forensic sample, approximately 56% of the sample experienced neurological or somatic injuries, whereas 42% experienced emotional problems (mostly mood and anxiety symptoms) as their primary concern.

The same validity scale exclusionary criteria used in the criminal sample were used to exclude claimants who exhibited non-responding or inconsistent responding on the MMPI-2. Three claimants were excluded. There were no significant differences between those excluded and those included in the civil forensic sample on any demographic.

Table 1
Description of the forensic samples

<table>
<thead>
<tr>
<th></th>
<th>Civil (n = 141)</th>
<th>Criminal (n = 127)</th>
<th>Statistical test</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Male</td>
<td>41.8</td>
<td>74.8</td>
<td>χ²(1, N = 268) = 29.70, p &lt; .001, η = .33</td>
</tr>
<tr>
<td>Mean age (S.D.)</td>
<td>43.3 (11.1)</td>
<td>34.3 (10.3)</td>
<td>t(265) = 6.85, p &lt; .001, d = .84</td>
</tr>
<tr>
<td>Mean education</td>
<td>13.8 (2.4)</td>
<td>11.7 (2.3)</td>
<td>t(262) = 7.60, p &lt; .001, d = .89</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Caucasian</td>
<td>66.0</td>
<td>59.8</td>
<td>χ²(1, N = 256) = 1.89, p = .17, η = .09</td>
</tr>
<tr>
<td>% African-American</td>
<td>12.8</td>
<td>36.2</td>
<td>χ²(1, N = 256) = 19.41, p &lt; .001, η = .28</td>
</tr>
<tr>
<td>% Latino/Latina</td>
<td>15.6</td>
<td>.8</td>
<td>χ²(1, N = 256) = 19.33, p &lt; .001, η = .28</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Married</td>
<td>51.8</td>
<td>11.8</td>
<td>χ²(1, N = 256) = 47.91, p &lt; .001, η = .43</td>
</tr>
<tr>
<td>% Single</td>
<td>22.0</td>
<td>54.3</td>
<td>χ²(1, N = 256) = 32.26, p &lt; .001, η = .36</td>
</tr>
<tr>
<td>% Divorced/separated</td>
<td>22.7</td>
<td>28.3</td>
<td>χ²(1, N = 256) = 1.37, p = .24, η = .07</td>
</tr>
</tbody>
</table>
or referral issue variable. With regard to SVTs, 85 claimants completed the TOMM and 86 completed the WMT. Thirty-two claimants took both symptom validity tests.

Several t-tests and $\chi^2$ were calculated to examine demographic differences between the two forensic samples and are presented in Table 1. The civil sample contained significantly more women and was significantly older and more educated than the criminal sample. The civil sample contained significantly fewer African-American and significantly more Latino/Latina participants than the criminal sample. Finally, the civil sample contained significantly more married and significantly fewer single participants. There were no significant differences between the two samples with regard to the percentage of Caucasian and divorced participants.

3.2. Instruments and materials

3.2.1. Minnesota multiphasic personality inventory-second edition (MMPI-2; Butcher et al., 2001)

The MMPI-2 is a self-report personality questionnaire comprised of 567 statements regarding symptoms, beliefs, and attitudes related to personality and psychopathology. The MMPI-2 manual (Butcher et al., 2001) and the RC scale monograph (Tellegen et al., 2003) provide extensive data regarding the psychometric characteristics of these scales in a variety of samples.

3.2.2. Test of memory malingering (TOMM)

The TOMM (Tombaugh, 1996) is a 50-item forced choice visual recognition test. During the learning phase of the test, 50 simple drawings of everyday objects (e.g., boat anchor, brief case, roller skates) are presented at a rate of 3 s per picture. The recognition trial involves showing the examinee 50 sets of pictures, which include one of the previous 50 pictures and a “foil” picture, which is a novel picture never presented again. The examinee must choose the picture that was shown during the learning phase. The process is repeated twice, and the examiner has the option of presenting a retention recognition trial 15 min after the completion of the second trial. The foil pictures during the recognition phases are not repeated, so that the examinee does not “learn” them. In a five experiment validation study, Rees, Tombaugh, Gansler, and Moczynski (1998) demonstrated that the TOMM had a sensitivity of 1.00 and a specificity of 1.00 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. Duncan (2005) found that the TOMM was not negatively impacted by psychotic symptomatology in a group of criminal forensic inpatients. Delain, Stafford, and Ben-Porath (2003) found that criminal defendants who failed the TOMM were more likely to report a previous head injury, to display marginal cooperation during testing, and to be diagnosed with antisocial personality disorder than those who passed the test.

3.2.3. Word memory test (WMT; Green, 2003; Green et al., 1996; Green & Astner, 1995)

The WMT is a forced-choice verbal recognition measure of effort that involves the presentation of 20 simple word pairs across two learning trials at a rate of one word pair per 2 s. After two learning phases of the test, the examinee completes an immediate 40-item forced-choice recognition test (IR), which involves the presentation of each word from the learning trial (target) a paired with a new word. The examinee must choose the target word in the pair from a new “foil” word that was not presented during the learning trials. The patient is presented with a similar format following a 30-min delay (Delayed Recognition Trial: DR). Research has found that the WMT is very sensitive to effort and insensitive to psychosocial variables, intelligence, psychopathology, and neurological impairment (Green, 2003). The WMT has been widely examined with patients who have sustained brain injuries (e.g., Green, Iverson, & Allen, 1999; Green et al., 2001), disability claimants with fibromyalgia (Gervais et al., 2001), and with schizophrenic patients (Gorissen, Sanz, & Schmand, 2005). The test however, has not been examined in a criminal forensic setting.

3.2.4. Record review forms

Two separate record review forms (criminal and civil forensic) were used to systematically code data from the evaluations. Both record review forms included basic demographics, referral information, forensic issue, and the clinician’s forensic opinion. In addition, the civil forensic record review coded information pertaining to the patient’s medical history and injury status. The criminal record review coded diagnostic information (i.e., DSM-IV Axis I and II diagnoses) and legal information (e.g., criminal history, current offense). Finally, the record reviews allowed for systematic coding of all psychological test scores. An independent reviewer who was trained to reliably code the information from the
reports calculated inter-rater reliability on 13% of the sample. Demographic variables demonstrated high inter-rater agreement, as measured by Kappa coefficients for dichotomous variables and interclass correlations for continuous variables, and ranged from .74 to 1.0, with an average inter-rater coefficient of .90. All interclass correlation coefficients for the criterion test scores were 1.0, demonstrating perfect agreement between reviewers.

3.3. Procedure

In both the criminal and civil forensic samples, patients/defendants were administered the MMPI-2 and symptom validity testing (WMT/TOMM) as part of their clinical evaluations. Data were extracted archivally using the record review forms.

4. Results

Cutoffs for the symptom validity tests used in the study were set in accordance with the tests’ manuals. Tombaugh (1996) applies a cutoff score of 90% on either Trial 2 or the Retention Trial of the TOMM, and Green (2003) uses a cutoff of 82.5% on either the Immediate Recall (IR) or Delayed Recall (DR) on the WMT as indicative of poor effort. Rather than examining the TOMM and WMT separately, participants were classified as either passed or failed on symptom validity testing depending on their performance on the TOMM or WMT. Among those who were administered both the TOMM and WMT, participants were placed in the “passed” SVT group if they passed both, or the “failed” SVT group if they failed at least one of the tests.4

Figs. 1 and 2 illustrate the mean profiles of individuals across the two settings who pass versus fail SVTs. These graphs show that on average, criminal defendants who fail SVTs tend to score highest on Scales 6, 8, RC6, and RC8, indicative of a psychotic presentation, whereas civil litigants who fail SVTs tend to score the highest on Scales 1, 2, 3, and RC1, which is consistent with a somatic complaints presentation.5

4.1. Hypothesis 1

To test our first hypothesis that criminal defendants who fail symptom validity tests will primarily produce marked elevations on MMPI-2 scales that measure psychotic symptoms, whereas civil forensic litigants who fail symptom validity testing will primarily produce marked elevations on MMPI-2 scales measuring somatic symptoms, we conducted a 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) × 2 (Scale RC1/RC8) ANOVA and found a significant three-way inter-

---

4 Discrepancies between the two symptom validity tests occurred in two of the sixteen criminal cases and nine of the thirty-two civil cases in which both tests were administered. In all of these cases, the TOMM was passed and the WMT was failed. Previous research has found a similar pattern between the TOMM and WMT. Green, Berendt, Mandel, and Allen (2000) attribute this discrepancy to the WMT having greater sensitivity to malingering than the TOMM.

5 Mean MMPI-2 profiles are limited in the amount of information that they convey given that it is possible that no single individual may match the average profile. Nevertheless, they can still be useful in depicting the average MMPI-2 performance across the various groups, providing a visual display of the type of complaints each group generally reported.
Fig. 2. Mean profiles for Restructured Clinical (RC) scales for criminal defendants and civil forensic litigants who pass and fail symptom validity tests. 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.

Fig. 3. Graphic display of the interaction terms for the 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVA predicting RC1 (upper left corner), RC8 (lower left corner), FBS (upper right corner), and Fp (lower right corner).

action effect for the RC Scales ($F_{[1,1248]} = 18.88, p < .001, \eta^2 = .07$). Follow-up 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVAs predicting scores on the individual somatic (RC1) and psychotic (RC8) scales were conducted. The interaction term for the prediction of RC1 was not statistically significant, $F_{(1, 237)} = .001, \ p = .980, \ \eta^2 = .000$. There was a significant main effect, however, for SVT status (i.e., pass or fail), for RC1 ($F_{[1,237]} = 79.98, \ p < .001, \ \eta^2 = .25$), in which those who failed SVTs scored higher on RC1 compared to those who passed SVTs, regardless of forensic setting. Fig. 3 (upper left corner) provides a graphic display of the non-significant follow-up univariate 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVA interaction terms for RC1. Table 2 (upper panel) shows the means

<table>
<thead>
<tr>
<th>Somatic scales</th>
<th>Criminal</th>
<th>Civil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass SVT ($n = 80$)</td>
<td>Fail SVT ($n = 34$)</td>
</tr>
<tr>
<td>RCd</td>
<td>RC1</td>
<td>66.8 (16.1)</td>
</tr>
<tr>
<td></td>
<td>RC1</td>
<td>68.3 (17.6)</td>
</tr>
<tr>
<td>RC1</td>
<td>63.7 (16.1)</td>
<td>88.4 (13.2)</td>
</tr>
<tr>
<td>RC8</td>
<td>FBS</td>
<td>64.4 (18.8)</td>
</tr>
</tbody>
</table>

Note: SVT: symptom validity test variable; RC: Restructured Clinical Scale; FBS: Fake Bad Scale and $F_p$: infrequency psychopathology.

* Linear T-scores were computed for the FBS based on the MMPI-2 normative sample.
and standard deviations for RC1 across all four conditions. We next examined the follow-up 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVAs predicting the psychotic scale (RC8). The interaction terms were statistically significant for RC8, $F(1, 248) = 14.80, p < .001, \eta^2 = .06$. Fig. 3 (lower left corner) provides a graphic display of the interaction terms. Table 2 (lower panel) shows the means and standard deviations for RC8 across all four conditions.

4.2. Hypothesis 2

To test our second hypothesis that criminal defendants who fail symptom validity testing will produce significant elevations on $F_P$, whereas the civil forensic patients who fail symptom validity testing will produce significant elevations on the Fake Bad Scale, we conducted a 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) × 2 (Validity Scale Type [$F_P$]/FBS) ANOVA. As expected, we found a significant three-way interaction effect, $F(1, 248) = 9.33, p < .001, \eta^2 = .04$. Follow-up 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVAs predicting scores on the FBS and $F_P$ scales were conducted. Similar to RC1, the interaction term for the prediction of the somatic validity scale (FBS) was not statistically significant, $F(1, 237) = .010, p = .922, \eta^2 = .000$. There was, however, a significant main effect for SVT status (i.e., pass or fail), for FBS ($F(1, 237) = 53.34, p < .001, \eta^2 = .18$), in that those who failed SVTs scored higher on the FBS compared to those who passed SVTs, regardless of forensic setting. Fig. 3 (upper right corner) illustrates the non-significant follow-up 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVA interaction term for FBS. Table 2 (upper panel) shows the means and standard deviations for the FBS across all four conditions. In a follow-up 2 (Pass/Fail SVT) × 2 (Civil/Criminal setting) ANOVAs predicting $F_P$, as expected, the interaction term was statistically significant, $F(1, 248) = 33.47, p < .001, \eta^2 = .12$. Fig. 3 (lower right corner) provides a graphic display of these interaction terms. Table 2 (lower panel) shows the means and standard deviations for $F_P$ across all four conditions.$^6$

5. Discussion

The purpose of the present study was to examine the relation of MMPI-2 scales and cognitive symptom validity test failure in two different types of forensic settings (civil versus criminal). Our findings indicated, as expected, that cognitive SVT failure was associated differentially with MMPI-2 symptom reporting indicators (RC1 versus RC8) and over-reporting measures (FBS versus $F_P$) as a function of setting. Specifically, we found that whereas SVT failure was associated with non-credible somatic complaining in both civil and criminal litigants, it was only associated with non-credible reporting of psychotic symptoms in criminal, but not in civil settings. This finding explains why the MMPI-2 infrequency scales such as $F_P$ are less effective than the FBS at identifying individuals found not to put forth their best effort on cognitive tasks in studies that have traditionally been conducted with civil, but not criminal litigants. Consistent with Larrabee’s (1998, 2003) observations, our findings show that the MMPI-2 infrequency scales do not tap non-credible somatic and cognitive symptoms presentation as well as the FBS.

Our results indicate that individuals who fail symptom validity tests, regardless of forensic context, produce marked elevations on an indicator of somatic complaining (RC1) and non-credible somatic complaining (FBS) on the MMPI-2. Whereas the Fake Bad Scale has been extensively examined in civil forensic settings, little is known about this scale in a criminal forensic setting (Greiffenstein, Fox, & Lees-Haley, 2007). The current study provides initial evidence that the scale identifies non-credible presentation of neurocognitive deficits across both civil and criminal forensic settings.

Although civil forensic patients who fail symptom validity tests tend to focus their self-reported symptoms on somatic and emotional complaints, the current results indicate that criminal defendants over-report more extreme forms of psychopathology as well when failing symptom validity tests. Our results also show that the previously reported limited association between MMPI-2 infrequency scales and cognitive symptom validity measures does not generalize to criminal forensic settings, supporting the notion that there are different demand characteristics of malingering and response bias in criminal settings, where trial incompetency and insanity are linked to more extreme forms of psychopathology. Given that the criminal defendants who failed symptom validity tests produced marked

---

$^6$ Our dependent variables were significantly associated with both gender and ethnic status, which are two demographic variables that are significantly different across the two settings. We therefore re-analyzed the data with analysis of covariance (ANCOVA) in which we partialled out the effect of both gender and ethnic status on the dependent variables. The pattern of results was identical to those reported in the text, suggesting that gender or ethnic differences does not account for the present findings.
elevations on somatic as well as psychotic scales, our results indicate that this group approached the testing in a less focused manner than the civil forensic litigants by exaggerating their symptoms across all three domains of potential response bias (somatic, cognitive, and psychological).

There are some limitations that need to be addressed in the present study. The two forensic samples differed with respect to gender and racial composition, level of education, marital status, and age. Thus, some may argue that differences across settings are due to demographics rather than the setting per se. However, differences on these demographic variables also reflect the typical characteristics of individuals assessed at these two types of forensic settings, with criminal defendants usually being younger, single, minority individuals with lower education levels than personal injury or disability clients. In other words, both samples were representative of their respective populations. Moreover, research has found that ethnicity and gender have little effect with regard to test bias on the MMPI-2 (Arbisi, Ben-Porath, & McNulty, 2002; Ben-Porath & Forbey, 2003; McNulty, Graham, Ben-Porath, & Stein, 1997). Also, as noted earlier, the pattern of findings was identical when the analyses were conducted after partialling out the effects of gender and race on the dependent variables.

To the extent that demographic characteristics actually account for some of the differences found across settings, this could explain, but does not alter the implications of, the observed interaction between forensic setting, performance on cognitive symptom validity testing, and scores on the MMPI-2 substantive and validity scales. Moreover, research with symptom validity measures indicates that demographic variables do not affect test scores. For example, Green and Flaro (2003) and Brockhaus and Merten (2004) found that age and intelligence did not impact WMT results. Delain et al. (2003) found no significant demographic differences between individuals passing or failing the TOMM with respect to age, gender, ethnicity, or years of education in a sample of criminal defendants. Similarly, Tombaugh (1997) found that the TOMM was insensitive to age and years of education in a sample of community dwelling adults.

In addition to demographic differences, the two forensic samples differed in symptom presentation, with approximately half of the civil sample presenting with somatic or neurological injuries, whereas the majority of the criminal defendants presented with substance abuse and personality psychopathology. Although it is possible that psychological status affected the results rather than the setting overall, it unlikely that this difference substantially detracts from our conclusions. The criminal sample primarily consists of externalizing forms of psychopathology, whereas the scales in question represented psychotic phenomena. Moreover, whereas previous literature has found that SVT scores are not significantly impacted by psychopathology, including psychosis (e.g., Duncan, 2005; Gorissen et al., 2005) and depression (e.g., Ashendorf, Constantinou, & McCaffrey, 2004; Rees, Tombaugh, & Boulay, 2001), future research should examine how differences in psychiatric presentation impact the relation between MMPI-2 scores and cognitive measures of response bias.

Follow-up research is needed to further explore the interaction between forensic assessment setting and malingering and its detection. Considerations of why demographic differences between participants in the two settings included in the present investigation do not alter the implications of our findings notwithstanding, a follow-up study with larger samples that would allow for separate analyses by gender, age, group, and race will help clarify the extent to which these demographics may explain our finding that the relation between symptoms validity testing and MMPI-2 scores is a function of the forensic evaluation context.

References


