

Further Validation of the MMPI-2 and MMPI-2-RF Response Bias Scale: Findings From Disability and Criminal Forensic Settings

Dustin B. Wygant
Eastern Kentucky University

Martin Sellbom
The University of Alabama

Roger O. Gervais
The University of Alberta

Yossef S. Ben-Porath
Kent State University

Kathleen P. Stafford
Cleveland VA Medical Center

David B. Freeman
Cal Psych FMT

Robert L. Heilbronner
Chicago Neuropsychology Group

The present study extends the validation of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) Response Bias Scale (RBS; R. O. Gervais, Y. S. Ben-Porath, D. B. Wygant, & P. Green, 2007) in separate forensic samples composed of disability claimants and criminal defendants. Using cognitive symptom validity tests as response bias indicators, the RBS exhibited large effect sizes (Cohen's $d_s = 1.24$ and 1.48) in detecting cognitive response bias in the disability and criminal forensic samples, respectively. The scale also added incremental prediction to the traditional MMPI-2 and the MMPI-2-RF overreporting validity scales in the disability sample and exhibited excellent specificity with acceptable sensitivity at cutoffs ranging from 90T to 120T. The results of this study indicate that the RBS can add uniquely to the existing MMPI-2 and MMPI-2-RF validity scales in detecting symptom exaggeration associated with cognitive response bias.

Keywords: Response Bias Scale, MMPI-2, MMPI-2-RF, MMPI-2 Restructured Form, malingering

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2001) and its original version have a long history of use in forensic settings and is the most widely administered ob-

jective personality test in forensic evaluations (Archer, Buffington-Vollum, Vauter Stredny, & Handel, 2006; Borum & Grisso, 1995; Lees-Haley, 1992), in part because of the test's extensive research base supporting its ability to detect response bias with its validity scales. The assessment of response bias is important in medicolegal and forensic evaluations, for which test users need to establish the validity of their interpretations. These forensic settings can provide motivation to distort or misrepresent one's clinical presentation. As such, defendants in civil litigation and prosecutors in criminal cases are interested in identifying when symptom exaggeration and fabrication occurs.

The traditional MMPI-2 validity scales (Variable Response Inconsistency [VRIN], True Response Inconsistency [TRIN], Infrequency [F], Back Infrequency [F_B], Infrequency Psychopathology [F_P], Lie [L], Correction [K], and Superlative Self-Presentation [S]) are well-established measures of response inconsistency and response bias involving overreporting or underreporting of psychopathology and emotional distress. A quasirare symptoms approach to response bias detection is used for the F scale. The quasirare detection strategy identifies symptoms that are rarely endorsed by individuals in normative, nonpatient populations. However, these symptoms are "quasirare" because they might also reflect genuine psychopathology among actual patients (Rogers, 2008). Although the F scale was originally developed to detect

This article was published Online First October 4, 2010.

Dustin B. Wygant, Department of Psychology, Eastern Kentucky University; Martin Sellbom, Department of Psychology, The University of Alabama; Roger O. Gervais, The University of Alberta, Edmonton, Alberta, Canada; Yossef S. Ben-Porath, Department of Psychology, Kent State University; Kathleen P. Stafford, Cleveland VA Medical Center, Cleveland, Ohio; David B. Freeman, Cal Psych FMT, Los Angeles, California; Robert L. Heilbronner, Chicago Neuropsychology Group, Chicago, Illinois.

Roger O. Gervais acknowledges being first author of the Response Bias Scale, from which he does not earn royalties. Yossef S. Ben-Porath is a paid consultant to the MMPI-2-RF publisher, the University of Minnesota Press, and test distributor, Pearson Assessments. He receives royalties on sales of MMPI-2-RF materials. Yossef S. Ben-Porath and Martin Sellbom receive research grants from the MMPI-2-RF publisher.

Portions of this study were presented at the 41st Annual Symposium of Recent Developments of the MMPI-2/MMPI-A, Minneapolis, Minnesota, May 2006.

Correspondence concerning this article should be addressed to Dustin B. Wygant, Department of Psychology, Eastern Kentucky University, 127 Cammack Building, 521 Lancaster Avenue, Richmond, KY 40475. E-mail: dustin.wygant@eku.edu

random responding, subsequent research indicated that the scale was often elevated among psychologically disturbed individuals and those attempting to present themselves as psychologically disturbed (Arbisi & Ben-Porath, 1995; Graham, 2006). The F_B scale is similar to F in its quasirare symptoms construction, in that it contained items infrequently endorsed in the MMPI-2 normative sample. A pure rare symptoms identification strategy is used in the F_P scale, containing items infrequently endorsed in a psychiatric sample (Arbisi & Ben-Porath, 1995). The “F-family” of validity scales (F , F_B , F_P) are used in detecting negative response bias associated with symptom exaggeration, particularly overreported symptoms of severe psychopathology. However, as discussed by Greiffenstein, Fox, and Lees-Haley (2007), the F-family validity scales were not originally developed for use within medico-legal settings, in which the authors contend that negative response bias is more likely to take the form of exaggerated somatic, cognitive, and emotional symptoms rather than psychotic symptoms. Moreover, Larrabee (2003) contended that personal injury claimants are more likely to present themselves as emotionally and physically injured rather than overtly psychotic, a notion that was discussed earlier by Lees-Haley (1989).

The above-noted limitations of the F-family validity scales in disability settings have led to the development of a number of other scales to detect exaggerated emotional complaints, such as the Malingered Depression (Md) Scale (Steffan, Clopton, & Morgan, 2003) and the Infrequency-Posttraumatic Stress Disorder scale (Fptsd; Elhai et al., 2002). The Infrequent Somatic Responses scale (Fs; Wygant, Ben-Porath, & Arbisi, 2004) was developed to measure exaggerated somatic complaints, whereas other scales and indices focused on the detection of exaggerated neurocognitive complaints, such as the Symptom Validity Scale¹ (FBS; Lees-Haley, English, & Glenn, 1991), the Meyers Validity Index (Meyers, Millis, & Volkert, 2002), the Henry-Heilbronner Index (HHI; Henry, Heilbronner, Mittenberg, & Enders, 2006), and the Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant, & Green, 2007), designed to be sensitive to overreporting in personal injury and medical disability contexts. The RBS is the focus of the present investigation. Although the scale was originally developed to detect negative response bias in forensic neuropsychological and disability settings, its ability to detect malingered neurocognitive deficits in other contexts has yet to be investigated empirically.

Development of the RBS

Gervais, Ben-Porath, et al. (2007) developed the RBS by examining differential item endorsement between individuals passing and failing cognitive symptom validity tests.² Using a series of multiple regressions on randomly split halves of a large sample of predominantly disability claimants, the authors identified 28 items that discriminated between individuals passing and failing several well-validated symptom validity tests, including the Word Memory Test (WMT; Green, 2003), Test of Memory Malingering (TOMM; Tombaugh, 1996), and the Computerized Assessment of Response Bias (Allen, Conder, Green, & Cox, 1997). The authors examined the RBS in a separate sample of more than 300 patients who completed the MMPI-2 and symptom validity testing as part of a disability evaluation and found that the scale demonstrated a large effect size (Cohen's $d = 0.92$) in discriminating between

individuals passing and failing the WMT and the Medical Symptom Validity Test (Green, 2004a). Furthermore, the RBS added incrementally to the F , F_P , and FBS scales in predicting WMT performance (incremental prediction ranging from an additional 6%–9% of variance). The RBS demonstrated sensitivity ranging from .34 (cutoff of 16) to .16 (cutoff of 18), with excellent specificity (.89 at a cutoff of 16 to .98 at a cutoff of 18).

In a follow-up study, Gervais, Ben-Porath, Wygant, and Green (2008) examined the relation between the RBS, the MMPI-2 F-family, and FBS and subjective self-reported memory complaints, as measured by the Memory Complaints Inventory (MCI; Green, 2004b). The RBS was more strongly correlated with all MCI scales than the MMPI-2 F-family and the FBS . For example, the correlation between the RBS and the mean score from the MCI was 0.69, compared with 0.32–0.56 for the MMPI-2 F-family and FBS . The authors also examined the association between objective memory performance, as measured by the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987), and the RBS. After controlling for effort by excluding individuals who failed one or more cognitive symptom validity tests, the correlations between the RBS and CVLT performance were nonsignificant and approached zero. Gervais, Ben-Porath, Wygant, and Sellbom (2010) showed similar results using the Restructured Form of the MMPI-2 (MMPI-2-RF; Ben-Porath & Tellegen, 2008). Indeed, the RBS added incrementally to the MMPI-2-RF validity scales and exhibited preferential ability to capture response bias while exhibiting little association to performance on an objective measure of memory performance after controlling for effort with cognitive SVTs.

Larrabee (2008) examined the diagnostic validity of the traditional MMPI-2 validity scales and the MMPI-2-RF validity scales, including the RBS, in discriminating between 41 malingering civil litigants and 54 nonmalingering patients with neurological disorders (mostly moderate to severe traumatic brain injury) and psychiatric illness (mostly major depression). FBS , RBS, and $FBS-r$ demonstrated the strongest effect sizes (Cohen's d) in discriminating between the malingering and nonmalingering groups ($d = 1.99, 1.91, \text{ and } 1.85$, respectively). Receiving operating characteristic curve (ROC) analysis produced excellent area under the curve (AUC) for each scale ($FBS = .917, RBS = .901, FBS-r = .900$). Regression analysis indicated that RBS was the best predictor of SVT performance ($R^2 = .436$) followed by $FBS-r$, the HHI (Henry et al., 2006), and the MMPI-2-RF Fs scale ($R^2 = .363, .270, \text{ and } .270$, respectively). Logistic regression illustrated that the combination of FBS and RBS accounted for the maximum variance in group membership and optimally classified 85.3% of subjects.

¹ The scale was formerly called “Fake Bad,” and the acronym FBS was retained to link the Symptom Validity Scale with the literature on its performance.

² Nelson, Sweet, and Heilbronner (2007) published empirical support of an earlier version of the RBS, which was based on a 39-item scale developed by Gervais (2005) and presented as a poster at the annual meeting of the National Academy of Neuropsychology. The published version of the RBS (Gervais, Ben-Porath, et al., 2007) contains 28 items and is the official version of the scale.

Smart et al. (2008) used classification tree methodology to examine the ability of the standard MMPI-2 validity and clinical scales, additional MMPI-2 validity scales and indices, and an earlier 41-item version of the RBS (Gervais, 2005) to predict cognitive effort in a sample of 307 individuals seen in either a secondary gain context ($n = 198$) or nonsecondary gain setting ($n = 109$). Their results revealed that the RBS was used first in classification of cognitive effort, followed by Hysteria.³ Although this analysis warrants replication with the final 28-item RBS, these results provide preliminary support for the construct validity of the RBS and its development objectives.

In a replication and extension of the original RBS validation study, Whitney, Davis, Shepard, and Herman (2008) examined the ability of the RBS and the other MMPI-2 overreporting validity scales (F , F_B , F_p , and FBS), the $Fptsd$ (Elhai et al., 2002), and the HHI (Henry et al., 2006) to predict failure on the TOMM in a sample of 46 outpatients referred for neuropsychological testing within a VA Medical Center. Whitney et al. found that the RBS had the largest effect size (Cohen's $d = 0.98$) in discriminating between groups who passed and failed the TOMM. Using hierarchical regression analyses, the authors examined the incremental validity of the RBS and concluded that the scale was superior to the MMPI-2 validity scales ($R^2 = .10-.20$, $p = .002-.009$), $Fptsd$ ($R^2 = .16$, $p = .005$), and to some extent the HHI in predicting symptom validity test failure within the study sample ($R^2 = .06$, $p = .069$).

The RBS was developed within a disability setting involving predominantly personal injury or medical disability claimants. The validation studies published to date support the RBS as an effective measure of cognitive symptom exaggeration and symptom validity test failure in disability settings. The effectiveness of the RBS in criminal forensic settings has not yet been examined. The goal of the present study was to further investigate the validity of the RBS in these settings. More specifically, we examined the RBS in two separate forensic settings, a sample of disability claimants and a sample of criminal defendants who completed psychological testing as part of a pretrial evaluation, with which the RBS has yet to be studied. Moreover, we examined the RBS in relation to both the MMPI-2 overreporting validity scales and their counterparts on the recently released MMPI-2-RF. It is hypothesized that the RBS will exhibit a stronger association with infrequency scales and markers of psychotic symptoms (i.e., RC8 [Aberrant Experiences]) in the criminal sample than in the disability sample. Previous research has found that individuals evaluated in a criminal forensic context exhibit broader overreporting of symptoms than civil forensic patients. Indeed, Duncan and Ausborn (2002) found a significant correlation between the Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994), a post hoc symptom validity measure created with the Wechsler Adult Intelligence Scale—Third Edition (Wechsler, 1997a) and the Wechsler Memory Scale—Third Edition (Wechsler, 1997b), and the MMPI-2 F scale in a sample of criminal forensic pretrial defendants. Moreover, Wygant et al. (2007) found that cognitive symptom validity test failure was related to exaggerated psychotic complaints on the MMPI-2 among criminal defendants. As a proxy for SVT performance, the RBS should exhibit a similar pattern of association with other MMPI-2 and MMPI-2-RF indicators of psychotic symptoms and exaggeration.

Method

Participants.

Criminal forensic. This present sample, previously examined in Wygant et al. (2007), consists of 127 criminal defendants referred for psychological evaluations to a certified forensic center providing psychological evaluations to five common pleas courts in northeastern Ohio. The defendants were predominantly men (75%) with a mean age of 34.1 years ($SD = 10.5$) and a mean education of 11.6 years ($SD = 2.4$). Regarding the ethnic and racial composition of the sample, the majority of the participants were Caucasian (60%); 36% were African American, and the race/ethnicity of the remaining 4% was unknown. Of the sample, 74% was referred for either a competency to stand trial or criminal responsibility (i.e., sanity) evaluation, and 26% were evaluated to determine suitability for a drug diversion program. 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 noncontent-based responding (Cannot Say ≥ 30), fixed inconsistent (TRIN T ≥ 80), and random responding (VRIN T ≥ 80). This resulted in the removal of 13 defendants who were significantly younger ($M = 28.4$, $SD = 6.6$) than included defendants ($M = 34.8$, $SD = 10.7$), $t(124) = 2.11$, $p = .037$. 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 TOMM, and 65 were administered the WMT. Sixteen defendants were administered both measures.

Disability. This sample, previously used in Wygant et al., 2007, and Wygant et al. 2009, consists of 141 personal injury or disability claimants referred by their insurance company, attorney, or worker's compensation for a psychological evaluation. The sample was predominantly women (58%), with a mean age of 43.3 years ($SD = 11.1$) and a mean education of 13.8 years ($SD = 2.4$). The ethnic and racial composition of the sample was predominantly Caucasian (66%), with 16% Latino/a, and 13% African American; the race/ethnicity of the remaining 5% was unknown. The majority of the sample claimed emotional disability due to work-related stress (42%), 33% experienced a minor head injury, and 16% experienced an orthopedic or musculoskeletal injury, and the remaining 9% experienced a nonhead injury neurological problem. Seventy percent of the claimants experienced work-related injuries, whereas 20% were involved in motor vehicle accidents; 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 disability sample, approximately 56% of the sample experienced neurological or somatic injuries, whereas 42% experienced emotional problems (mostly

³ Although the Smart et al. (2008) study used an earlier version of the RBS, the correlation between the 41-item RBS and the official 28-item RBS is .86, which suggests that the authors' findings should transfer well to the final scale. Nevertheless, this assumption needs to be tested empirically.

mood and anxiety symptoms) as their primary concern; diagnostic information was not available for the remaining 2% of the disability sample. The same validity scale exclusionary criteria used in the criminal sample were used to exclude claimants who exhibited nonresponding or inconsistent responding on the MMPI-2, which resulted in the removal of three claimants. There were no significant differences between those excluded and included in the disability sample on any demographic or referral issue variable. With regard to SVT administration, 85 claimants completed the TOMM, and 86 completed the WMT. Thirty-two individuals were administered both symptom validity tests.

Measures.

The MMPI-2 (Butcher et al., 2001). The MMPI-2 is a self-report personality questionnaire composed of 567 statements regarding symptoms, beliefs, and attitudes related to personality and psychopathology. The manual provides extensive information regarding the basic psychometric properties of the test. For this study, the MMPI-2 Validity Scales (i.e., F, F_B, F_P, FBS, and RBS) and Restructured Clinical (RC) scales (Tellegen et al., 2003) were used.

The MMPI-2-RF (Ben-Porath & Tellegen, 2008). 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. The test's four overreporting validity scales (Infrequent Responses, F-r; Infrequent Psychopathology Responses, Fp-r; Fs; and FBS-r) as well as the RC scales were examined for comparative purposes with the RBS.⁴ Please refer to the Appendix for a scoring guide to the MMPI-2-RF RBS.

The WMT (Green, 2003; Green, Allen, & Astner, 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 (Immediate Recall), which involves the presentation of each word from the learning trial (target) within a pair of two words. The examinee must choose the word from the target word in the pair from a novel "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). 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 examined in a wide range of settings, including those of patients who have sustained brain injuries (e.g., Green, Iverson, & Allen, 1999; Green, Rohling, Lees-Haley, & Allen, 2001), disability claimants with fibromyalgia (Gervais et al., 2001), personal injury litigants and workers' compensation claimants (Stevens, Friedel, Mehrem, & Merten, 2008), patients diagnosed with schizophrenia (Gorissen, Sanz, & Schmand, 2005), and criminal defendants (Ardolf, Denney, & Houston, 2008).

The TOMM (Tombaugh, 1996). The TOMM is a 50-item 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 pic-

tures, which include one of the previous 50 pictures and a "foil" picture, which is a new picture. The examinee must choose the picture that was shown during the learning phase. The process is repeated, 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 .96 in comparing a group of brain-injured patients instructed to malingering 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, display marginal cooperation during testing, and to be diagnosed with antisocial personality disorder more than those who passed the test.

Procedure. Claimants/defendants were administered the MMPI-2 and symptom validity testing (WMT/TOMM) as part of their clinical evaluations.⁵ Data were extracted archivally. Cutoffs for determining failure on the WMT and the TOMM were set in accordance with their respective test manuals. Individuals who failed any SVT were assigned to the "fail SVT" group. Claimants/defendants who passed all SVTs administered to them were assigned to the "pass SVT" group. The SVT failure rate was 30% in both the criminal and disability groups. This figure is consistent with the 30%–35% SVT failure rates reported in previous studies using personal injury, medical disability, and criminal forensic samples (Gervais, Rohling, Green, & Ford, 2004; Gierok, Dickson, & Cole, 2005; Greve et al., 2006). The formation of the pass and fail SVT groups in this manner was pragmatic, in that given the archival nature of the data, the authors were not able to control the fact that some individuals were only administered one SVT, whereas others were administered both the TOMM and WMT. Moreover, designation of the "fail" group with failure on only one SVT is supported by previous research (see Bianchini, Mathias, & Greve, 2001; Green, 2007), which suggests that cognitive SVTs are designed to be highly specific, but with relatively low sensitivity. Consequently, failing an SVT is a more significant finding than passing an SVT.

Among participants who completed both the TOMM and WMT, discrepant results between the measures occurred in two of the 16 criminal defendants and nine of the 32 civil litigants. In all of these cases, the TOMM was passed, and the WMT was failed. Previous research has found a similar pattern between the two measures (Gervais, Rohling, Green, & Ford, 2004), which was attributed to the WMT's greater sensitivity to malingering than the TOMM.

⁴ The MMPI-2-RF overreporting validity scales were previously examined in the disability sample and presented in Wygant et al. (2009). They are only presented in this study for comparative purposes with the RBS.

⁵ *Sensitivity* refers to the proportion of individuals correctly classified as having a condition, whereas *specificity* refers to the proportion of individuals correctly classified as not having the condition.

Results

Zero-order correlations. To examine the type of symptoms associated with the RBS in the respective settings, we calculated zero-order correlations between this scale and the standard MMPI-2/MMPI-2-RF validity and the RC (Tellegen et al., 2003) scales. These correlations are shown in Table 1. As expected, the RBS was significantly correlated with all of the validity scales and most of the RC scales across both samples. Examination of the correlations between the RBS and the RC scales illustrates that scores on the scales are associated with a diverse set of symptom claims, with the exception of externalizing behaviors, as indicated by nonsignificant correlations with RC4 (Antisocial Behavior) and RC9 (Hypomanic Activation). We calculated Fisher's z tests to determine whether any differences in the correlation pattern between the two samples were statistically significant. As expected, the RBS was correlated significantly higher with F, F_B , F_P , F-r, Fp-r, and RC8 (Aberrant Experiences) in the criminal sample, suggesting that higher RBS scores are related to overreporting of psychopathology symptoms to a greater extent in criminal rather than in disability settings.

Validity scale group differences. T tests and effect sizes for the MMPI-2/MMPI-2-RF validity scales comparing individuals who passed and failed symptom validity measures were calculated in both samples and are reported in Tables 2 and 3. As seen in Table 2, F_P ($d = 1.65$) and F ($d = 1.61$) were associated with the highest effect sizes in predicting SVT failure in the criminal forensic group, but they were likely not meaningfully higher than RBS ($d = 1.48$), F-r ($d = 1.48$), and Fp-r ($d = 1.46$). In the disability group (see Table 3), the RBS was associated with the largest effect size ($d = 1.24$) in discriminating between pass/fail SVT subgroups, but likely not meaningfully better than FBS ($d = 1.18$), Fs ($d = 1.14$), and FBS-r ($d = 1.13$).

Incremental validity. We next calculated hierarchical logistic regression analyses to examine whether the RBS added incremental predictive utility to the MMPI-2 and MMPI-2-RF standard validity scales, respectively, in predicting cognitive symptom validity test failure across the two samples. In these analyses, we simultaneously entered all standard MMPI-2 or MMPI-2-RF validity scales in the first block and then the RBS in the second block. Table 4 presents the results of these analyses in the criminal forensic sample. These results showed that the RBS did not add incremental predictive utility to either the MMPI-2 or MMPI-2-RF set of standard overreporting validity scales in the criminal sample. Table 5 displayed these analyses in the disability sample. Here, as expected, RBS added incremental predictive utility to both the MMPI-2 and MMPI-2-RF overreporting validity scales.

Classification accuracy. Classification accuracies of the RBS in differentiating symptom validity test failure in the two respective samples were examined using ROC curves, which are calculated on the basis of a function of sensitivity and $1 - \text{specificity}$.⁶ Predictive performance was assessed using the area under the ROC curve (AUC). We then selected optimal cut-off scores in terms of sensitivity and specificity for further analyses. 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 for the scale. 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.

The RBS produced an AUC of .850 ($SE = .037$; 95% CI = .777, .923) in the criminal forensic sample, compared with an AUC of .810 ($SE = .037$; 95% CI = .737, .882) in the disability sample, indicating excellent power to predict SVT failure.

Tables 6 and 7 display the classification accuracy of the RBS in predicting SVT failure in the criminal and disability group, respectively. In the criminal sample, T -score cutoffs between 100 and 120 produced optimal balance of sensitivity and specificity. At a cutoff score of 100, specificity was .89 and sensitivity was .59, resulting in a PPP of .69. Increasing the cutoff to 120 resulted in a false-positive rate of only 1%, but with a sensitivity of .24 as a result. The overall hit rate ranged from .71 to .80. Given that the RBS did not add incrementally to other measures in the criminal sample, we did not examine combinations of this scale and other scales in the dichotomous prediction of SVT performance.

In the disability sample (see Table 7), RBS cutoff scores higher than 90 were associated with excellent specificity, ranging from .91 to 1.00, but substantially lower sensitivity ranging from .38 to .02. PPP ranged from .50 to 1.00, with an overall hit rate between .70 and .75. Table 7 also provides classification accuracy for several combinations of RBS and MMPI-2/MMPI-2-RF validity scales. For instance, when using the MMPI-2, applying a cutoff score of 80 on the RBS in combination with a cutoff of 80 on FBS, sensitivity drops from .69 to .62, but specificity increases from .70 to .81, as opposed to only using a cutoff of 80 on the RBS. Using a combination of the RBS at 80 and F, F_B , or F_P at 80 produced a sensitivity of .67 with a specificity of .81. Raising the RBS cutoff to 90 in combination with a cutoff of 80 on FBS resulted in a specificity of .91 (with a sensitivity of .36). Using RBS at a cutoff of 90 in combination of a cutoff of 80 on F, F_B , or F_P resulted in a specificity of .97 (with a sensitivity of .31).

When using the MMPI-2-RF, applying a combination of RBS at 80 with Fs or FBS-r at 80 yielded a specificity of .81 and a sensitivity of .67. Using the RBS at 90 with Fs or FBS-r at 80 resulted in a specificity of .93 and a sensitivity of .38. Applying the same cutoff for RBS (90) with a cutoff of 80 for F-r or Fp-r yielded the same classification results.

Discussion

The goal of the present study was to extend the initial validation studies of the RBS by examining its ability to predict symptom

⁶ To explore whether gender differences, rather than the nature of the setting, could potentially explain the differences in results across the two settings, we conducted a two-way multivariate analysis of variance, with group membership and gender as independent variables and all validity scales as dependent variables. We conducted these analyses separately for MMPI-2 and MMPI-2-RF as well as for each setting. There was only one significant interaction—in the disability sample, there was a Gender \times SVT Failure interaction for MMPI-2-RF validity scales. Follow-up univariate two-way analyses of variance (ANOVAs) revealed that this interaction was specific to Fs. The effect size for this Gender \times SVT Group interaction was small ($\eta^2 = .035$), and ANOVAs separate by gender revealed that Fs remained significantly higher in the fail SVT group relative to the pass SVT group for both genders. Thus, it is unlikely that gender effects explain these results.

Table 1
Correlations With the Response Bias Scale (RBS)

Variable	F	F _B	F _P	FBS	F-r	Fp-r	Fs	FBS-r	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
Criminal	.88*	.87*	.70*	.77*	.91*	.78*	.76*	.80*	.78*	.77*	.72*	.36*	.27	.73*	.76*	.83*	.10
Civil	.74*	.72*	.42*	.71*	.82*	.59*	.63*	.72*	.77*	.77*	.69*	.32*	.08	.50*	.65*	.60*	.15
z	3.32*	3.32*	3.28**	1.04	2.89**	2.87**	1.99	1.49	0.20	0.00	0.47	0.35	1.54	2.96	1.72	3.86*	0.40

Note. Raw scores are used for correlations. F = Infrequency; F_B = Back Infrequency; F_P = Infrequency Psychopathology; FBS = Symptom Validity; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RC = Restructured Clinical; 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; z = Fischer's z test calculated to compare the difference in correlation magnitude across the two samples.
* p ≤ .001. ** p = .001.

validity test failure in two clinical samples representing distinct forensic settings. Although these samples are separate from the sample used to create the RBS, the disability sample is comparable to the RBS development sample in terms of its composition of disability claimants.

The results in the disability sample are comparable to the initial RBS validation sample presented in Gervais, Ben-Porath, et al. (2007) in which the RBS outperformed the F-Family and FBS in identifying SVT performance. The criminal forensic sample represents a qualitatively distinct clinical population from the development or initial validation samples from which the RBS was originally developed. Nevertheless, the RBS was associated with a large effect size in the criminal sample (*d* = 1.48) in discriminating between defendants passing or failing SVTs. Moreover, the RBS was correlated significantly higher with F, F_B, F_P, F-r, Fp-r, and RC8 (Aberrant Experiences) in the criminal sample versus the civil sample, suggesting that higher RBS scores are associated with overreporting of psychopathology symptoms to a greater extent in criminal, rather than in civil, forensic settings. This finding is not surprising when one considers that response bias in most criminal forensic settings (particularly competency to stand trial and criminal responsibility evaluations) is associated with exaggerated psychotic symptoms conceptually associated with “incompetency” and “insanity” (Melton, Petrila, Poyhress, & Slobogin, 2007). Further review of the correlations between the RBS and the scales

examined in the study (see Table 1) suggest that the RBS is associated with both psychotic symptoms and psychotic overreporting. This might be attributable to the notion that the individuals who score high on this scale also endorse numerous symptom complaints on the substantive clinically oriented scales on the MMPI-2/MMPI-2-RF. SVT failure, and by extension, RBS performance, appears to tap into the larger dimension of symptom exaggeration as an underlying attitude. Evaluatees can sometimes be quite focused and specific in what they overreport, but in other cases, they exaggerate in many or on all dimensions of functioning. Therefore, it is not surprising that the RBS is associated with psychotic symptom reporting (particularly in criminal settings), in the same way that there is often a substantial correlation between F and F-r with measures of psychopathology. However, the specificity of the RBS with regards to feigned psychiatric symptoms needs to be further examined in psychiatric examples comprising individuals who have genuine psychosis and limited to no motivation for symptom overreporting. The issue of whether validity scales on the MMPI-2 should be specifically developed to detect particular types of symptom feigning is not a new one. Indeed, earlier studies, such as that of Bagby, Rogers, and Buis (1994), found that the Gough Dissimulation Index, the Gough Dissimulation Scale (Ds; Gough, 1954), and the Wiener (1948) Obvious-Subtle Index (O-S), exhibited utility at detecting feigned psychiatric symptoms. Moreover, Rogers, Bagby, and Chakraborty

Table 2
Comparisons Between Those Who Passed Symptom Validity Testing (SVT) (*n* = 80) and Those With Failed SVT Results (*n* = 34) in the Criminal Forensic Sample

Scale	Passed SVT		Failed any SVT		<i>t</i> (112)	<i>p</i>	Effect size	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			Cohen's <i>d</i>	95% CI
RBS	70.3	21.7	101.2	19.0	7.2	<.001	1.48	1.02, 1.91
F	75.2	22.6	109.2	17.0	7.9	<.001	1.61	1.15, 2.05
F _B	79.8	25.5	111.9	13.7	6.9	<.001	1.42	0.97, 1.85
F _P	64.4	18.8	97.2	22.1	8.1	<.001	1.65	1.19, 2.10
FBS	68.3	17.6	84.8	15.2	4.8	<.001	0.97	0.55, 1.39
F-r	81.7	25.1	114.3	11.5	7.2	<.001	1.48	1.03, 1.90
Fp-r	68.2	23.4	102.2	23.2	7.1	<.001	1.46	1.00, 1.89
Fs	71.2	23.7	100.5	19.7	6.3	<.001	1.30	0.85, 1.72
FBS-r	67.1	16.8	84.7	13.5	5.4	<.001	1.11	0.67, 1.53

Note. All scores are *T* scores. CI = confidence interval; RBS = Response Bias Scale; F = Infrequency; F_B = Back Infrequency; F_P = Infrequency Psychopathology; FBS = Symptom Validity; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity.

Table 3
Comparisons Between Those Who Passed Symptom Validity Testing (SVT) ($n = 96$) and Those With Failed SVT Results ($n = 42$) in the Disability Sample

Scale	Passed SVT		Failed any SVT		$t(136)$	p	Effect size	
	M	SD	M	SD			Cohen's d	95% CI
RBS	67.1	17.5	88.1	15.7	6.7	<.001	1.24	.84–1.62
F	57.1	14.3	70.7	19.2	4.6	<.001	.85	.47–1.22
F _B	56.8	18.4	74.2	21.8	4.8	<.001	.89	.51–1.27
F _P	50.9	10.4	58.0	14.8	3.2	<.001	.60	.22–.96
FBS	69.9	16.6	88.1	12.0	6.4	<.001	1.18	.79–1.57
F-r	65.0	18.8	85.7	21.8	5.7	<.001	1.05	.66–1.42
F _P -r	52.1	12.6	60.5	14.9	3.4	.001	.63	.26–1.00
F _S	59.0	17.3	80.0	20.6	6.2	<.001	1.14	.75–1.52
FBS-r	69.2	15.8	85.9	11.9	6.1	<.001	1.13	.74–1.51

Note. All scores are T scores. CI = confidence interval; RBS = Response Bias Scale; F = Infrequency; F_B = Back Infrequency; F_P = Infrequency Psychopathology; FBS = Symptom Validity; F-r = Infrequent Responses; F_P-r = Infrequent Psychopathology Responses; F_S = Infrequent Somatic Responses; FBS-r = Symptom Validity.

(1993) found that the Ds-r2 and O-S were effective at detecting feigned schizophrenia on the MMPI-2, even when dissimulators were coached. Additionally, validity scales such as the Fptsd (Elhai et al., 2002) and Md (Steffan et al., 2003) were developed to identify specifically feigned symptom presentations. However, subsequent research by Bagby and colleagues (e.g., Bagby et al., 1997; Bury and Bagby, 2002) found that the family of F scales (i.e., F, F_B, F_P) were the most effective scales at detecting specific forms of feigned psychopathology, such as schizophrenia, depression, and posttraumatic stress disorder. Moreover, Marshall and Bagby (2006) found that the family of F scales outperformed the Fptsd scale in detecting feigned PTSD, and Bagby, Marshall, and Bacchioni (2005) found that the Md scale did not provide much clinical utility over the family of F scales in detecting feigned depression.

Hierarchical logistic regression analyses confirmed the incremental validity of the RBS by demonstrating that this scale added to the standard MMPI-2 and MMPI-2-RF validity scales in predicting SVT failure in the disability sample. However, the RBS did

not add incremental predictive utility to either set of scales in the criminal sample. Thus, the RBS, in isolation, performs approximately as well as other validity scales in differentiating SVT performance, but it is unlikely to provide any additional information to clinicians above and beyond what is already derived from the standard set of validity scales on either version of the test in criminal forensic settings.

The results of the standard MMPI-2 and MMPI-2-RF infrequency validity scales in the criminal sample supports Larrabee's (2003) contention that symptom exaggeration in the disability setting is more likely to take the form of overreported physical symptoms consistent with individuals presenting themselves as injured, whereas in the criminal forensic arena, defendants are more likely to overreport severe psychopathology in order to present themselves as being thought disordered. Nevertheless, the large effect size associated with the RBS in the criminal forensic group was not meaningfully smaller than that of the infrequency scales and indicates that the scale is particularly sensitive to exaggeration of cognitive dysfunction as a component of the

Table 4
Regression Analyses in the Criminal Forensic Sample ($N = 114$)

Step	Parameters					
	Model fit $\chi^2(df)$	w''	$\Delta\chi^2(df)$	$\Delta w''$	R^2	ΔR^2
MMPI-2						
F, F _B , F _P , FBS	54.57 (4)**	.69			.38	
RBS	54.65 (5)**	.69	.08 (1)	.03	.38	.00
MMPI-2-RF						
F-r, F _P -r, F _S , FBS-r	55.85 (4)**	.70			.39	
RBS	56.52 (5)**	.70	.67 (1)	.08	.39	.00

Note. The Cox and Snell R^2 estimation, derived from SPSS, was used for logistic regression. w'' = effect size for chi-square statistic; $\Delta w''$ = effect size for change in chi-square statistic; MMPI-2 = Minnesota Multiphasic Personality Inventory-2; F = Infrequency; F_B = Back Infrequency; F_P = Infrequency Psychopathology; FBS = Symptom Validity; RBS = Response Bias Scale; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2 Restructured Form; F-r = Infrequent Responses; F_P-r = Infrequent Psychopathology Responses; F_S = Infrequent Somatic Responses; and FBS-r = Symptom Validity.
** $p < .01$.

Table 5
Regression Analyses in the Disability Sample (N = 138)

Step	Parameters					
	Model fit $\chi^2(df)$	w''	$\Delta\chi^2(df)$	$\Delta w''$	R^2	ΔR^2
MMPI-2						
F, F _B , F _P , F _{BS}	40.74 (4)**	.54			.26	
RBS	46.00 (4)**	.58	5.26 (1)*	.20	.28	.02
MMPI-2-RF						
F-r, F _P -r, F _S , F _{BS} -r	42.52 (4)**	.56			.27	
RBS	47.46 (5)**	.59	4.94 (1)*	.19	.29	.02

Note. The Cox and Snell R^2 estimation, derived from SPSS, was used for logistic regression. w'' = effect size for chi-square statistic; $\Delta w''$ = effect size for change in chi-square statistic; MMPI-2 = Minnesota Multiphasic Personality Inventory-2; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2 Restructured Form; F-r = Infrequent Responses; F_P-r = Infrequent Psychopathology Responses; F_S = Infrequent Somatic Responses; F_{BS}-r = Symptom Validity.
** $p < .01$.

exaggerated psychopathology that this clinical group can present with during evaluations. Although the RBS was initially designed to measure symptom overreporting associated with poor performance on cognitive symptom validity tests (Gervais, Ben-Porath, et al., 2007), results of the present study provides some preliminary evidence that the scales might also be associated with psychopathological overreporting. However, additional research examining the scale's ability to detect feigned psychopathology will need to be conducted before any conclusion on this matter is reached. A recent study by Lange, Sullivan, and Scott (2010) found that the RBS was among the best MMPI-2 validity scales at detecting feigned symptoms of PTSD and depression in an analogue sample of college students, outperforming the diagnosis-specific validity scales F_{ptsd} and Md. However, the findings were limited by rather small groups of students instructed to feign psychopathology ($n = 15$ for PTSD and 14 for depression), and the control group included 20 students instructed to respond honestly; thus, a comparison group of clinical patients was not used in the study. Additional simulation studies in which genuine patient control groups are used will be needed before definitive conclusions can be reached regarding this matter. Moreover, it is recommended that future validation research examine the extent to which the RBS is asso-

ciated with psychopathological symptom exaggeration on established response bias measures such as the Structured Interview of Reported Symptoms (Rogers, Bagby, & Dickens, 1992) or the Miller Forensic Assessment of Symptoms Test (Miller, 2001).

Examination of the classification accuracy statistics for the two samples identified different optimal cutoffs for predicting SVT failure in the criminal and disability samples. Cutoffs of 90–100 were associated with specificity ranging from .91 to .98 and PPP ranging from .64 to .83 in the disability sample. This range is similar to the original RBS validation study (Gervais, Ben-Porath, et al., 2007). This range of cutoffs minimizes false-positive errors, albeit at the expense of sensitivity (.24–.38). In the criminal sample, cutoffs of 100–20 produced acceptable to excellent spec-

Table 6
Response Bias Scale (RBS) Classification Accuracy in the Criminal Forensic Sample (N = 114)

Scale cutoff (T)	SENS	SPEC	PPP	NPP	HR
RBS 120	.24	.99	.89	.75	.76
RBS 110	.44	.95	.79	.80	.80
RBS 100	.59	.89	.69	.84	.80
RBS 95	.68	.84	.64	.86	.79
RBS 90	.71	.76	.56	.86	.75
RBS 85	.79	.74	.56	.89	.75
RBS 80	.85	.65	.51	.91	.71

Note. Base rate = .30. SENS = sensitivity; SPEC = specificity; PPP = positive predictive power; NPP = negative predictive power; HR = overall hit rate.

Table 7
Response Bias Scale (RBS) Classification Accuracy in the Disability Sample (N = 138)

Cutoff (T)	SENS	SPEC	PPP	NPP	HR
RBS 120	.02	1.00	1.00	.70	.70
RBS 110	.12	.98	.71	.72	.72
RBS 100	.24	.98	.83	.75	.75
RBS 95	.31	.95	.72	.76	.75
RBS 90	.38	.91	.64	.77	.75
RBS 85	.48	.84	.57	.79	.73
RBS 80	.69	.70	.50	.84	.70
RBS 80 & F/F _B /F _P 80	.36	.91	.63	.76	.74
RBS 80 & F _S 80	.62	.81	.59	.83	.75
RBS 80 & F-r, F _P -r 80	.43	.84	.55	.77	.72
RBS 80 & F _S /F _{BS} -r 80	.67	.81	.61	.85	.77
RBS 90 & F/F _B /F _P 80	.31	.97	.81	.76	.77
RBS 90 & F _S 80	.36	.91	.63	.76	.74
RBS 90 & F-r/F _P -r 80	.38	.93	.70	.77	.76
RBS 90 & F _S /F _{BS} -r 80	.38	.93	.70	.77	.76

Note. Base rate = .30. SENS = sensitivity; SPEC = specificity; PPP = positive predictive power; NPP = negative predictive power; HR = overall hit rate; F = Infrequency; F_B = Back Infrequency; F_P = Infrequency Psychopathology; F_S = Symptom Validity; F-r = Infrequent Responses; F_P-r = Infrequent Psychopathology Responses; F_S = Infrequent Somatic Complaints; F_{BS}-r = Symptom Validity.

ificity (.89–.99), substantially stronger sensitivity at .59–.74, and good PPP of .69–.89. These findings suggest that the use of different RBS cutoff scores in civil and criminal forensic settings may be warranted. However, regardless of setting, clinicians should be aware that due to the sensitivity rates given the base rates for these samples, a substantial majority of respondents exhibiting cognitive response bias will not be identified as such. This might be attributable to the different modes of response bias (i.e., self-report vs. performance based); however, this needs to be further explored. Results by Demakis, Gervais, and Rohling (2008), who examined PTSD claimants on a battery of psychological and neuropsychological measures, found a differential predictive ability between SVTs and performance on neuropsychological testing and self-report validity scales and measures of psychopathology.

The classification results also suggest that the combination of cutoffs between the RBS and other validity scales can reduce false-positive classifications in disability settings. This is consistent with Gervais, Lees-Haley, and Ben-Porath (2007), who examined combinations of the RBS with FBS/FBS-r and Fs in a sample of nonhead injury disability claimants and found that SVT failure rates increased when individuals were elevated on multiple MMPI-2/MMPI-2-RF validity scales. In the present study, the combination of RBS with other MMPI-2 and MMPI-2-RF validity scales increased the specificity by incorporating the other scales' focus on exaggerated psychopathology and somatic complaints. This is consistent with the guidelines presented by Gervais et al. (2008) in interpreting the RBS as a measure of exaggerated self-reported memory complaints with other validity scales.

Regardless of whether SVTs are passed or failed in an individual case, it is important to recall that elevated RBS scores have been found to be associated with increased subjective memory/cognitive complaints (Gervais et al., 2008). The RBS, by virtue of its design and validation, is in a unique position among the other MMPI-2/MMPI-2-RF response bias indicators to provide compelling evidence of cognitive exaggeration, at least in disability settings. It is also quite possible that the cognitive exaggeration detected by the RBS might not be accompanied by SVT failure. As noted by Nelson, Sweet, Berry, Bryant, and Granacher (2007), "Individual litigants may demonstrate extremely poor cognitive effort in the presence of unremarkable MMPI-2 profiles, or conversely, exhibit extreme elevations on multiple MMPI-2 validity scales and demonstrate excellent cognitive effort" (p. 446). Thus, in cases in which SVT performance suggests good cognitive effort, elevated RBS scores nonetheless signal the possibility of exaggerated cognitive symptoms, which warrants further examination. It is important to emphasize that neither cognitive SVTs nor self-report measures are 100% accurate. Consequently, it is important to use various measures (across various domains of measurement) to assess response bias.

As a cautionary note, although RBS scores have been found to be associated with performance on SVTs, it still remains imperative that clinicians continue to use SVTs as a direct assessment of feigned cognitive dysfunction, a point that is strongly supported in position statements by both the National Academy of Neuropsychology (Bush et al., 2005) and the American Academy of Clinical Neuropsychology (Heilbronner et al., 2009). Both of these position statements, however, along with proposed diagnostic standards for assessing malingered cognitive dysfunction (Slick, Sherman, &

Iverson, 1999) and malingered pain-related disability (Bianchini, Greve, & Glynn, 2005) indicate the value of incorporating evidence from self-report measures such as the MMPI-2 along with performance-based symptom validity tests when assessing response bias. The data from this study and others on the RBS (e.g., Gervais, Ben-Porath, et al., 2007, 2008, 2010; Whitney et al., 2008) suggest that the scale can provide complimentary evidence in support of findings from SVTs when evaluating the complex nature of response bias. Moreover, as Gervais, Ben-Porath, et al. (2007) pointed out, the RBS might also serve as a retrospective proxy for performance on cognitive SVTs in situations in which formal measures of cognitive response bias were not administered, or in which relatively insensitive SVTs were used.

Similar to our results in the disability sample, Gervais et al. (2008) found that the RBS had its largest correlation with RCd and RC1, reflecting an association with emotional and somatic distress. Although the results of this study and previous examinations of RBS have found association between the scale and measures of emotional symptom presentation on the MMPI-2, as discussed earlier, additional research is needed before conclusions about the scale's ability to assess emotional response bias can be reached. Gervais and his colleagues (2008) suggested that an elevation on the RBS might be tapping "emotional" exaggeration in cases in which the individual has passed symptom validity tests. Future research should attempt to tease these concepts apart empirically. One suggestion for potentially investigating this issue would be comparing scores on measures of emotional distress, such as the Beck Depression Inventory-II (Beck, Steer, Brown, 1996), between individuals who have passed symptom validity tests and produce elevated scores on the RBS with individuals who have passed symptom validity tests and produce nonelevated scores on the RBS.

The results of this study support and extend the initial validation of the RBS. The disability sample produced results that are remarkably consistent with the original RBS validation sample described by Gervais, Ben-Porath, et al. (2007). Results for the criminal forensic sample revealed that the RBS, at a slightly higher cutoff, demonstrated substantially better sensitivity at acceptable levels of specificity than in the civil sample. Sensitivity rates in both samples, however, were generally not much better than base rates. Findings of this study suggest that the RBS can be a valuable adjunct to the existing MMPI-2/MMPI-2-RF validity scales in predicting effort test failure associated with exaggerated cognitive symptoms in disability settings. Future research should explore the properties of the scale with respect to structured malingering criteria such as the Malingered Neurocognitive Dysfunction (Slick et al., 1999) and the Malingered Pain-Related Disability (Bianchini et al., 2005). Moreover, research investigating the relationship between the RBS and other MMPI-2/MMPI-2-RF indicators of negative response bias with client report of inaccurate clinical history or dramatic discrepancy between observed behavior and test-taking style is also warranted.

There were several limitations to this study that need to be acknowledged. First, portions of the data used in the present study were presented in previous studies (Wygant et al., 2009, 2007). The continued use of these two data sets potentially increases the risk that sampling errors contributing to the findings can be exaggerated in the research literature. However, the RBS has not been previously examined in these samples.

Although SVT's test scores tend to be associated with acceptable reliability estimates given their forced-choice nature (Bianchini et al., 2001), we did not have access to individual test items due to the archival nature of the database and, thus, were unable to estimate these properties here. Nonetheless, given the magnitude of effect sizes reported in the study, it is unlikely that unreliability of test scores would significantly have attenuated the findings.

One might describe the archival data used in the present study as samples of convenience, particularly the quarter of the criminal sample that were referred for drug diversion evaluations. Although there is little inherent motivation for this group of criminal defendants to feign neurocognitive impairment, their inclusion in the study serves two purposes. First, from a psychometric standpoint, their inclusion with the competency/sanity evaluatees increases the variance of RBS scores and SVT performance in the analyses, as there is less likelihood that these individuals would score high on RBS or exhibit response bias on SVTs. Psychometrically, this increase in variance enhances our correlational analyses. Second, from a more pragmatic clinical position, one should not artificially bias the sample of investigation toward only those with substantial motivation to feign symptoms. Indeed, the motivation for symptom misrepresentation is present across most forensic settings, and one can never know whether response bias occurs unless it is formally assessed. In this regard, the RBS might serve as a useful screening measure in settings in which one would not inherently expect to find feigned cognitive deficits. Using the scale in these settings where cognitive symptom validity tests are typically not administered routinely, clinicians can screen for cognitive response bias, which could then be followed up with direct measures of cognitive effort.

Moreover, the results may have been influenced by differences in symptom presentation between the two forensic samples, with approximately half 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 psychopathology affected the results rather than the forensic context overall, it is unlikely that this difference substantially detracts from our conclusions. Although 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 and MMPI-2-RF scores and cognitive measures of response bias. Finally, there were significantly different gender ratios in the two samples, suggesting that there may be potential gender effects rather than differences in psychopathology or evaluation context explaining the results. Although we mentioned previously that exploratory analyses indicate that gender effects are unlikely (see Footnote 6), the specific samples sizes for some cells were very low. Therefore, future investigations should more carefully study potential gender effects for the RBS and other MMPI-2/MMPI-2-RF across different contexts.

In summary, the RBS was shown to be associated with poor performance on cognitive symptom validity tests in both criminal and disability settings. Although detection of cognitive response bias should include symptom validity testing, SVTs (like all psychological tests) are susceptible to coaching and also have limited

sensitivity. The RBS can provide important corroborating evidence of response bias, or in some cases, the scale might provide indications of possible response bias (e.g., rule out coaching of SVTs) in cases in which SVTs are passed. Future studies are needed to examine the validity of the RBS in evaluating somatic and psychopathological symptoms, as well as examining the general performance of the scale in nonforensic and clinical settings.

References

- Allen, L., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB 97 manual for the computerized assessment of response bias*. Durham, NC: CogniSyst, Inc.
- 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., Vauter Stredny, R., & 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. (2008). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21, 899–916.
- Ashendorf, L., Constantinou, M., & McCaffrey, R. J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology*, 19, 125–130.
- Bagby, R. M., Marshall, M. B., & Bacchiochi, J. R. (2005). The validity and clinical utility of the MMPI-2 Malingering Depression scale. *Journal of Personality Disorders*, 85, 304–311.
- 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.
- Bagby, R. M., Rogers, R., Buis, T., Nicholson, R. A., Cameron, S. L., Rector, N. A., . . . Seeman, M. V. (1997). Detecting feigned depression and schizophrenia on the MMPI-2. *Journal of Personality Assessment*, 68, 650–664.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for Beck Depression Inventory II (BDI-II)*. San Antonio, TX: Psychological Corporation.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *The Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, 5, 404–417.
- Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist*, 15, 19–45.
- Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice*, 26, 465–473.
- Bury, A. S., & Bagby, R. M. (2002). The detection of feigned uncoached and coached posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment*, 14, 472–484.
- Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., . . . Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, 20, 419–426.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2: Manual for administration and scoring* (Rev. ed.). Minneapolis: University of Minnesota Press.
- Delain, S. L., Stafford, K. P., & Ben-Porath, Y. S. (2003). Use of the TOMM in a criminal court forensic assessment setting. *Assessment*, 10, 370–381.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California*

- Verbal Learning Test (CVLT) manual*. San Antonio, TX: The Psychological Corporation.
- Demakis, G. J., Gervais, R. O., & Rohling, R. L. (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist, 22*, 879–895.
- Duncan, A. (2005). The impact of cognitive and psychiatric impairment of psychotic disorders on the Test of Memory Malingering (TOMM). *Assessment, 12*, 123–129.
- Duncan, S. A., & Ausborn, D. L. (2002). The use of reliable digits to detect malingering in a criminal forensic pretrial population. *Assessment, 9*, 56–61.
- Elhai, J. D., Ruggiero, K. J., Frueh, B. C., Beckham, J. C., Gold, P. B., & Feldman, M. E. (2002). The Infrequency-Posttraumatic Stress Disorder Scale (Fptsd) for the MMPI-2: Development and initial validation with veterans presenting with combat-related PTSD. *Journal of Personality Assessment, 79*, 531–549.
- Gervais, R. (2005, April). *Development of an empirically derived response bias scale for the MMPI-2*. Paper presented at the annual MMPI-2 Symposium and Workshops, Ft. Lauderdale, FL.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14*, 196–208.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *The Clinical Neuropsychologist, 22*, 1061–1079.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2-RF over-reporting scales and RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology, 25*, 274–284.
- Gervais, R. O., Lees-Haley, P. R., & Ben-Porath, Y. S. (2007, November). *Predicting SVT performance with the MMPI-2-RF, FBS-r, RBS, and Fs scales*. Poster presented at the 27th Annual Meeting of the National Academy of Neuropsychology, Scottsdale, AZ.
- Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19*, 475–487.
- Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology, 28*, 1892–1899.
- Gierok, S. D., Dickson, A. L., & Cole, J. A. (2005). Performance of forensic and non-forensic adult psychiatric inpatients on the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 20*, 755–760.
- Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research, 78*, 199–208.
- Gough, H. G. (1954). Some common misperceptions about neuroticism. *Journal of Consulting Psychology, 18*, 287–292.
- Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th ed.). New York, NY: Oxford University Press.
- Green, P. (2003). *Green's Word Memory Test for Windows: User's manual*. Edmonton, Canada: Green's Publishing.
- Green, P. (2004a). *Green's Medical Symptom Validity Test (MSVT): User's manual*. Edmonton, Canada: Green's Publishing.
- Green, P. (2004b). *Memory Complaints Inventory*. Edmonton, Canada: Green's Publishing.
- Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America, 18*, 43–68.
- Green, P., Allen, L., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms, US Version 1.1*. Durham, NC: CogniSyst, Inc.
- Green, P., & Astner, K. (1995). *Manual for the Oral Word Memory Test*. Durham, NC: CogniSyst, Inc.
- Green, P., Iverson, G., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13*, 813–819.
- Green, P., Rohling, M., Lees-Haley, P., & Allen, M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*, 1045–1060.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 218–224.
- Greiffenstein, M. F., Fox, D. D., & Lees-Haley, P. R. (2007). The MMPI-2 Fake Bad Scale in detection of noncredible brain injury claims. In K. B. Boone (Ed.), *Detection of noncredible cognitive performance* (pp. 210–235). New York, NY: Guilford Press.
- Greve, K. W., Bianchini, K. J., Black, W. F., Heiny, M. Y., Love, J. M., Swift, D. A., & Ciota, M. (2006). Classification accuracy of the Test of Memory Malingering in persons reporting exposure to environmental and industrial toxins: Results of a known-groups analysis. *Archives of Clinical Neuropsychology, 21*, 439–448.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.
- Henry, G. K., Heilbronner, R. L., Mittenberg, W., & Enders, C. (2006). The Henry-Heilbronner Index: A 15-item empirically derived MMPI-2 subscale for identifying probable malingering in personal injury litigants and disability claimants. *The Clinical Neuropsychologist, 20*, 786–797.
- Lange, R. T., Sullivan, K. A., & Scott, C. (2010). Comparison of MMPI-2 and PAI validity indicators to detect feigned depression and PTSD symptom reporting. *Psychiatry Research, 176*, 229–235.
- Larrabee, G. J. (2003). Exaggerated MMPI-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of Clinical Neuropsychology, 18*, 673–686.
- Larrabee, G. (2008, February). *Evaluation of new MMPI-2 validity scales in malingering civil litigants and non-malingering clinical patients*. Paper presented at the 36th annual meeting of the International Neuropsychological Society, Waikoloa, Hawaii.
- Lees-Haley, P. (1989). MMPI F and F-K scales: Questionable indices of malingering. *American Journal of Forensic Psychology, 7*, 81–84.
- Lees-Haley, P. R. (1992). Psychodiagnostic test usage by forensic psychologists. *American Journal of Forensic Psychology, 10*, 25–30.
- 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.
- Marshall, M. B., & Bagby, R. M. (2006). The incremental validity and clinical utility of the MMPI-2 Infrequency Posttraumatic Stress Disorder scale. *Assessment, 13*, 417–429.
- Melton, G. B., Petrila, J., Poyhress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts* (3rd ed.). New York, NY: Guilford Press.
- Meyers, J. E., Millis, S. R., & Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology, 17*, 157–169.
- Miller, H. A. (2001). *M-FAST: Miller Forensic Assessment of Symptoms Test professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Nelson, N. W., Sweet, J. J., Berry, D. T., 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.
- Nelson, N. W., Sweet, J. J., & Heilbronner, R. L. (2007). Examination of the new MMPI-2 Response Bias Scale (Gervais): Relationship with MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 29*, 67–72.
- Rees, L. M., Tombaugh, T. N., & Boulay, L. (2001). Depression and the

- Test of Memory Malingering. *Archives of Clinical Neuropsychology*, *16*, 501–506.
- Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, *10*, 10–20.
- Rogers, R. (2008). *Clinical Assessment of Malingering and Deception* (3rd ed.). New York, NY: Guilford Press.
- Rogers, R., Bagby, R. M., & Chakraborty, D. (1993). Faking schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment*, *60*, 215–226.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Slick, D. J., Sherman, E., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.
- Smart, C. M., Nelson, N. W., Sweet, J. J., Bryant, F. B., Berry, D. T., Granacher, R. P., & Heilbronner, R. L. (2008). Use of MMPI-2 to predict cognitive effort: A hierarchically optimal classification tree analysis. *Journal of the International Neuropsychological Society*, *14*, 842–852.
- Steffan, J. S., Clopton, J. R., & Morgan, R. D. (2003). An MMPI-2 scale to detect malingered depression (Md scale). *Assessment*, *10*, 382–392.
- Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, *157*, 191–200.
- Tellegen, A., & Ben-Porath, Y. S. (2008). *The Minnesota Multiphasic Personality Inventory-2 Restructured Form: Technical manual*. Minneapolis: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI-2 Restructured Clinical (RC) Scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.
- Tombaugh, T. N. (1996). *Test of Memory Malingering*. Toronto, ON: MultiHealth Systems.
- Wechsler, D. A. (1997a). *Wechsler Adult Intelligence Scale-Third edition*. New York, NY: Psychological Corporation.
- Wechsler, D. A. (1997b). *Wechsler Memory Scale-Third edition*. New York, NY: Psychological Corporation.
- Whitney, K. A., Davis, J. J., Shepard, P. H., & Herman, S. M. (2008). Utility of the Response Bias Scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology*, *23*, 777–786.
- Wiener, D. N. (1948). Subtle and obvious keys for the MMPI. *Journal of Consulting Psychology*, *12*, 164–170.
- 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 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.

Appendix

Response Bias Scale Items for the MMPI-2-RF and Nongendered T Scores

RBS raw	T score	RBS raw	T score
0	29	12	80
1	33	13	84
2	38	14	88
3	42	15	92
4	46	16	97
5	50	17	101
6	54	18	105
7	59	19	109
8	63	20	114
9	67	21	118
10	71	22	120
11	76	23+	120

Note. Source: MMPI[®]-2 (Minnesota Multiphasic Personality Inventory[®]-2) Manual for Administration, Scoring, and Interpretation, Revised Edition. Copyright © 2001 by the Regents of the University of Minnesota. Used by permission of the University of Minnesota Press. All rights reserved. “MMPI” and “Minnesota Multiphasic Personality Inventory” are trademarks owned by the Regents of the University of Minnesota. T scores derived from the MMPI-2 Minnesota normative sample. True: 6, 24, 26, 31, 68, 74, 79, 92, 101, 106, 120, 132, 136, 137, 159, 242, 252, 268, 273. False: 11, 21, 53, 59, 125, 131, 156, 219, 325. $M = 4.95$, $SD = 2.37$.

Received June 17, 2009
 Revision received April 19, 2010
 Accepted April 19, 2010 ■