Framing the Results: Assessment of Response Bias Through Select Self-Report Measures in Psychological Injury Evaluations

Chelsea E. Sleep · Jade A. Petty · Dustin B. Wygant

Abstract Forensic psychological injury evaluations require extensive consideration of malingering and response bias. These evaluations are complicated in that they require assessment of both symptoms and symptom exaggeration across multiple domains of functioning (e.g., psychological, somatic, neurocognitive). Self-report measures are routinely utilized in psychological injury evaluations both for their economy of use and their broad coverage of symptoms. The following article reviews the use of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2001), MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011), and Personality Assessment Inventory (PAI; Morey, 2007). We will briefly review each instrument, focusing on their over-reporting validity indicators, and offer some general comments about their use in forensic psychological injury evaluations.

Keywords Forensic assessment · Psychological injury · MMPI-2 · MMPI-2-RF · PAI

The use of psychological testing in medicolegal settings has increased over the past 30 years, as emotional damages have been increasingly considered in civil suits (Butcher & Miller, 1999). In addition to its capturing a wide range of psychological functioning, self-report measures are particularly effective in assessing various forms of response bias (Rogers & Granacher, 2011). This is important in medicolegal and forensic assessments, where psychologists must establish the validity of their test interpretations. Civil litigation suits can be highly conflictual and financial damages can motivate litigants to over-report symptoms or over-state the nature of their impairment. Consequently, defendants (including insurance companies) are motivated to establish the veracity of plaintiff damages or dysfunction before awarding financial compensation (Miller, Sadoff, & Dattilio, 2011). The adversarial nature of the courts requires a level of objectivity in establishing dysfunction that is often not required in treatment settings. Several professional organizations have taken the lead in establishing practice guidelines that dictate the necessity of including response bias measures in forensically oriented evaluations, including the National Academy of Neuropsychology (Bush et al., 2005), American Academy of Clinical Neuropsychology (Heilbronner et al., 2009), and more recently, the Association for Scientific Advancement in Psychological Injury and Law (Bush, Heilbronner, & Ruff, 2014). Given the inherent potential for self-reported symptoms to be misrepresented, it is important that self-report assessment measures include ways to measure response bias. While some have challenged the need for self-report measures to include embedded validity indicators (see McGrath, Mitchell, Kim, & Hough, 2010), other studies have demonstrated that response bias (particularly over-reporting) can attenuate the associations between a measure and relevant criteria (Burchett & Ben-Porath, 2010; Wiggins, Wygant, Hoelzle, & Gervais, 2012). Moreover, clinical experience will often illustrate that symptom feigning on psychological testing will often be corroborated with other evidence (Burchett & Bagby, 2014; Rogers, 2008). For instance, elevated self-report validity scales can be accompanied by dramatic reporting of symptoms during a clinical interview.

While a number of brief, symptom-focused measures are useful in psychological injury evaluations [e.g., Trauma Symptom Inventory-2 (TSI-2; Briere, 2010)], the current review will focus on three broadband, omnibus measures of personality and psychopathology that are particularly popular in forensic settings and psychological injury evaluations

C. E. Sleep · J. A. Petty · D. B. Wygant (✉)
Department of Psychology, Eastern Kentucky University, 521
Lancaster Ave., Richmond, KY 40475, USA
e-mail: dustin.wygant@eku.edu

28

Psychol. Inj. and Law (2015) 8:27–39

because of their well-established validity indicators: the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2001), MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011), and Personality Assessment Inventory (PAI; Morey, 2007). These three measures are broad in their coverage of psychological dysfunction and include well-established embedded validity indicators. Rather than providing an extensive discussion of test development and interpretation for the MMPI-2, MMPI-2-RF, and PAI, which is covered in detail by the tests' respective manuals and interpretive texts (e.g., Ben-Porath, 2012; Blais, Baity, & Hopwood, 2010; Graham, 2011; Morey, 2003), the current review will instead provide a brief description of the measures and instead discuss their over-reporting scales and particular utility in the assessment of response bias in psychological injury evaluations.

Before discussing specific over-reporting indicators, it is worth noting that most of the response bias indicators embedded in the MMPI-2, MMPI-2-RF, and PAI are premised on the notion that individuals misrepresenting their symptoms lack a nuanced understanding of psychopathology. One particular approach employed by several self-report validity scales, termed *Rare Symptoms*, incorporates the frequency of psychopathology symptoms (Rogers & Bender, 2012). Many individuals feigning symptoms endorse those items that are rarely experienced among patients with genuine illness. Those lacking a sophisticated understanding of bona fide symptoms of mental illness may mistakenly over-endorse symptoms that actually occur rarely among genuine patients. Another response bias detection strategy, termed the *Erroneous Stereotypes* approach, assumes that dissimulators may endorse erroneous conceptions of psychopathology (Rogers, 2008). An example of this would be to assume that every person suffering from schizophrenia experiences command hallucinations. Utilizing this approach, validity scales include items judged by professionals to be suggestive of psychopathology and maladjustment, but which were not typically endorsed by actual patients, giving them the opportunity to endorse erroneous stereotypes of psychopathology. Finally, other validity indicators on the MMPI-2-RF and PAI were developed through empirical means.

Minnesota Multiphasic Personality Inventory-2

The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) has played a significant role in the history of personality assessment. Developed in the 1940s, the MMPI set the standard for personality assessment measures for decades, enjoying wide use in clinical, medical, and forensic settings among others. However, by the 1970s, it was evident that the test's normative sample was outdated and many of the items needed reassessment, alteration, or removal

to cater to a new era of social norms and psychological disorders (Sellbom & Anderson, 2013). As a result, the MMPI re-standardization project was launched in the 1980s, and the MMPI-2 was released in 1989. The MMPI-2 norms attempted to match the 1980 US census. As indicated earlier, it is beyond the scope of this review to cover the reliability and validity of the MMPI-2, which is unmatched by current psychological measures in terms of its empirical examination. The interested reader is directed to recent interpretive texts such as Graham (2011) and Greene (2011) for a comprehensive review.

MMPI-2 Over-Reporting Indicators

The MMPI-2 utilizes four standard validity scales to identify over-reported symptoms. Hoelzle, Nelson, and Arbisi (2012) provide an excellent review of the strategies employed by these scales, emphasizing ways in which the MMPI-2 and MMPI-2-RF validity scales complement one another. The first of the MMPI-2's over-reporting indicators is the Infrequency (F) scale. This 60-item scale was originally developed for the MMPI to identify atypical responding suggestive of random responding, but it was later discovered that elevations on this scale could also indicate over-reporting (Graham, 2011). The F scale items all appear in the initial 370 items of the test and reflect heterogeneous content. In particular, this scale includes rare psychopathological symptomology endorsed by less than 10 % of the original MMPI normative sample.[1] A number of F items appear in scales measuring psychopathology. Therefore, psychologically disturbed individuals tend to endorse a large number of these items. Individuals attempting to portray themselves in an unrealistic negative light are also inclined to endorse these items (Graham, 2011). Elevations of this scale are thought to be suggestive of a response style reflective of an attempt to portray oneself as more maladjusted than they are, a "cry for help," random responding or true psychopathology. Given the various causes for elevation on F, elevations on this scale should not be interpreted in isolation, but rather in relation the other validity scales, particularly the Back Infrequency (F_B) and Infrequency Psychopathology (Fp) scales (Graham, 2011).

Since the F scale has been found to reflect both a genuine psychopathology and an exaggerated response style, scores on this scale can have different meanings in clinical (inpatient and outpatient) and a nonclinical settings. According to interpretive guidelines, scores above 100T in inpatient, 90T in outpatient, and 80T for nonclinical settings (Butcher et al., 2001). In particular, nonclinical scores in this range are

---

[1] Note that when the original MMPI F items were transferred to the MMPI-2, some of these items had endorsement rates above 10 % for the MMPI-2 normative sample. Thus, while these items may have been infrequently endorsed by the original MMPI normative standard (of less than 10 %), this may not have been the case for every one of these items in relation to the MMPI-2 norms (Graham, 2011).

interpreted as an attempt to present oneself in an unrealistically negative light, whereas for inpatient and outpatient groups, it could be a consequence of genuine psychopathology. It is not uncommon to see elevations exceeding 100T in forensic settings, particularly in settings with high rates of posttraumatic stress disorder (PTSD) or malingering (Elhai et al., 2002). Thus, elevations on F must be considered in relation to other MMPI-2 over-reporting scales, particularly Fp, as well as indicators of non-content-based invalid responding (i.e., Variable Response Inconsistency and True Response Inconsistency).

Given that the F scale items occur early in the instrument, the scale is unable to determine the validity of the responses appearing latter half. Subsequently, the F_B scale was developed. It contains 40 items occurring between items 281 and 555. It is conceptually similar to the F scale, with fewer than 10 % of the MMPI-2 normative sample endorsing these items in the scored direction. Moreover, the F and F_B scales are highly correlated (Butcher et al., 2001), but items on the F_B are more reflective of atypical emotional distress and tend to be more elevated in inpatient psychiatric samples (Arbisi & Ben-Porath, 1995). Elevations on the F_B scale (110T in clinical settings; 90T in nonclinical; Butcher et al., 2001) indicate the test taker responded to the second half of the measure in an invalid manner. The MMPI-2 manual suggests when F_B is significantly elevated and at least 30T-score points higher than F, the clinician can conclude that the test taker may have changed his or her approach, and scales utilizing items in the final half of the test (e.g., restructured clinical and content scales) should be interpreted with caution. Similar to the F scale, elevations on F_B are suggestive of either exaggerated or fabricated psychological symptomology or a genuine psychopathology; therefore, it too should not be interpreted in isolation (Graham, 2011).

Since the F scale is often elevated among patients with severe psychopathology, Arbisi and Ben-Porath (1995) developed the Fp scale to supplement it. Fp was developed to detect exaggerated psychological symptoms. Additionally, it is unique to the other F-family scales in that its 27 items were rarely endorsed (by less than 20 %) in two large psychiatric inpatient samples, as well as the MMPI-2 normative sample. Fp has also been shown to add incrementally to the F scale in differentiating between psychiatric inpatients and over-reporting (Arbisi & Ben-Porath, 1995; Bagby et al., 1997; Graham, 2011). Therefore, if an elevation is observed on F or F_B, clinicians should review Fp to differentiate between genuine psychopathology and noncredible symptom exaggeration (Arbisi & Ben-Porath, 1995; Butcher et al., 2001; Steffan, Morgan, Lee, & Sellbom, 2010). Lastly, since this scale was developed in relation to psychiatric samples, it is less sensitive to bona fide psychopathology and does not require adjusted interpretive guidelines dependent on the population for which it is used, although Nichols (2011)

cautioned that Fp can be elevated when all of the family enmity items are endorsed even if no other Fp items are endorsed. According to the MMPI-2 manual, elevations (>100 T) are suggestive of an exaggerated response style and are likely invalid and should not be interpreted (Butcher et al., 2001). Scores ranging from 70 to 99T are indicative of probable exaggeration or a "cry for help." Nonetheless, the profile should still be considered valid and interpretable (Graham, 2011).

Originally termed the *Fake Bad Scale*, the Symptom Validity Scale (FBS) was developed to identify inflated emotional distress among personal injury claimants (Lees-Haley, English, & Glenn, 1991). [2] It includes 43 rationally selected items that were thought to be reflective of exaggerated post-injury emotional distress, while also minimizing any preexisting psychopathology. In particular, these items were generated from a review of items commonly endorsed in civil forensic settings as well as from the scale authors' experience within that setting (Hoelzle et al., 2012). It has also been suggested that in addition to emotional distress and physical functioning, this scale assesses cognitive complaints, sleep disturbances, and energy as well as morality (Henry, Heilbronner, Mittenberg, Enders, & Stanczak, 2008).

FBS was added to the standard MMPI-2 scoring in 2006, and a subsequent test monograph was published (Ben-Porath, Graham, & Tellegen, 2009). The scale's addition did not come without controversy or criticism. In particular, it has been argued that it incorrectly classifies females and individuals with genuine medical problems as malingering (Butcher, 2010; Butcher, Arbisi, Atlis, & McNulty, 2003; Butcher, Gass, Cumella, Kally, & Williams, 2008; Dean et al., 2008; Gass, Williams, Cumella, Butcher, & Kally, 2010; Williams, Butcher, Gass, Cumella, & Kally, 2009). However, Ben-Porath et al. (2009) demonstrated that when the recommended interpretive cutoff scores (FBS≥100T) were employed, few people with bona fide medical or neurological disorders were mistakenly recognized as noncredible responders. Moreover, in a sample of medical patients with sleep disturbances, FBS scores did not appear to be influenced by the presence of medical impairments (Greiffenstein, 2010). This scale has also exhibited utility in a preliminary study distinguishing noncredible responding from individuals with brain injuries as well as from those with conversion disorder (Peck et al., 2013). In addition, Lee, Graham, Sellbom, and Gervais (2012) found that when proper cutoff scores are used, the scale is equally valid for both men and women and can satisfactorily distinguish persons simulating cognitive disorders from those with genuine cognitive disorders. Wygant et al. (2007) also explored the association between MMPI-2 scales and

---

[2] Although the scale is now referred to as the *Symptom Validity Scale* rather than its original *Fake Bad Scale* title, the shortened FBS is still used to denote this scale.

30

Psychol. Inj. and Law (2015) 8:27–39

cognitive performance validity tests (PVT[3]) in two samples comprising criminal defendants who were assessed for drug dependence, competency to stand trial, and criminal responsibility, and civil litigants were assessed for personal injury and disability claims. Results indicated that FBS scores were related to PVT failure in both groups; however, scores on the Fp scale were only linked with PVT failure in criminal defendants. Therefore, the FBS scale is thought to be a useful indicator of exaggerated somatic and cognitive complaints in both civil and forensic assessment and perhaps more so than the F-family scales. Research has further supported that in cognitive evaluations, FBS elevations were associated with suspected malingering or poor effort in traumatic brain injury (TBI) litigants, such that FBS elevations were inversely related to TBI severity (Thomas & Youngjohn, 2009). In their meta-analysis, Nelson, Hoelzle, Sweet, Arbisi, and Demakis (2010) concluded that studies examining FBS since their previous meta-analysis (Nelson, Sweet, & Demakis, 2006) have more than doubled. In addition, the cumulative literature suggests that FBS has a large composite effect and is stable in detecting noncredible responding, particularly in psychological injury evaluations, and in some cases better than the F-family scales.

MMPI-2 and Psychological Injury and Related Evaluations

The MMPI-2 has a long history of use in forensic settings, including psychological injury evaluations. Given the complexity of these types of evaluations, where the evaluee can present with psychological, as well as somatic and cognitive symptoms, it is important that instruments used the evaluation can assess response bias in these areas. The MMPI-2 has been examined in relation to the various potential forms of response bias, including feigned emotional distress (Crawford, Greene, Dupart, Bongar, & Childs, 2006), including PTSD (discussed later in details), as well as criteria for Malingered Neurocognitive Dysfunction (MND; Slick, Sherman, & Iverson, 1999) and Malingered Pain-Related Disability (MPRD; Bianchini, Greve, & Glynn, 2005). Greve, Bianchini, Love, Brennan, and Heinly (2006) examined the MMPI-2 validity scales in a sample of traumatic brain injury and general clinical neuropsychological patients. These authors found that the MMPI-2 validity scales (particularly FBS and F_B) exhibited good classification in discriminating MND from non-malingering patients. Utilizing both simulated malingerers and bona fide chronic pain patients, Bianchini, Etherton, Greve, Heinly, and Meyers (2008) found that the MMPI-2 validity scales (again, especially FBS and F_B) were

able to accurately differentiate malingerers from non-malingerers based on the MPRD criteria.

The MMPI-2 is well-known for its ability to assess feigned symptoms of PTSD (Demakis & Elhai, 2011). Specifically, Fp has shown an accuracy rate in the range of 80–90 % in a variety of studies when using a cutoff of 85T (Arbisi, Ben-Porath, & McNulty, 2006; Efendov, Sellbom, & Bagby, 2008; Marshall & Bagby, 2006). Arbisi et al. (2006) assessed the efficacy of the MMPI-2's validity scales in their ability detect feigned PTSD in compensation- and pension-seeking veterans. The veterans were randomly assigned to either feign PTSD or respond honestly. The results revealed that although all of the infrequency scales were able to identify noncredible symptom exaggeration, Fp had the best overall hit rate. Additionally, Tolin et al. (2010) assessed the F-family scales' ability to detect symptom exaggeration in a sample of veterans examined for service-connected disability for PTSD. Using a mixed group validation approach with varying base rate estimates, the authors found that F, F_B, Fp, and FBS were all able to accurately detect an exaggerated response style among veterans undergoing VA PTSD evaluations with adequate sensitivity, specificity, and efficiency across various base rate estimates. In another study, Efendov, Sellbom and Bagby (2008) examined the ability of the MMPI-2 to detect feigned PTSD. Using a sample of remitted trauma victims, who completed the MMPI-2 and Trauma Symptom Inventory (TSI; Briere, 1995), these authors assigned the sample retake both measures and fake symptoms of PTSD. Half of the sample was provided with coaching regarding the validity scales on both measures, the other half was not provided any additional information. These groups were compared with workplace injury claimants with PTSD. They found that F, F_B, and Fp were able to distinguish both the coached and non-coached samples from the current PTSD claimants. FBS was only able to distinguish the non-coached sample from the PTSD claimants. Furthermore, of the F-family scales, outperformed FBS and the Atypical Response Scale on the TSI in classifying feigned PTSD (Efendov, Sellbom & Bagby, 2008).

MMPI-2 Restructured Form

The Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011) was developed to capture the clinical substance of the MMPI-2 item pool while employing more sophisticated scale construction to enhance the measure's psychometric properties. Since the clinical scales were left nearly identical when the MMPI was revised in 2001, many of the same psychometric concerns (e.g., poor discriminant validity, high intercorrelations) that were raised with the original MMPI were present with the MMPI-2. Consequently, Tellegen et al. (2003) sought to restructure the Clinical scales in order to

---

[3] Larrabee (2012) advocated for referring to symptom validity tests (SVTs) as performance validity tests (PVTs) to clarify the extent to which an examinee's test performance is accurately reflecting their actual level of ability.

improve their psychometric functioning by isolating the effects of general emotional distress, which they termed "demoralization," upon each of the clinical scales. The Restructured Clinical (RC) scales were successful in improving the psychometric functioning of the MMPI-2 clinical scales. These scales eventually served as the primary clinical indicators on the MMPI-2 Restructured Form (MMPI-2-RF). These nine scales were augmented by 33 other substantive measures and nine validity indicators to reliably and validly measure a full range of constructs assessed by the MMPI-2. While the MMPI-2-RF validity scales are generally similar to their MMPI-2 counterparts, there are some differences between the two sets of scales. Perhaps the most significant change is the decreased item overlap among the scales, with the exception of FBS-r and RBS (Hoelzle et al., 2012).

MMPI-2-RF Over-Reporting Indicators

There are five over-reporting validity scales on the MMPI-2-RF. The Infrequent Responses (F-r) scale is the counterpart of the MMPI-2's F scale and contains 32 items distributed throughout the measure. Like F, it was developed as a measure of general over-reporting by including items rarely endorsed (≤10 %) in the normative sample (Tellegen & Ben-Porath, 2008/2011). Elevations on the F-r scale may indicate over-reporting, inconsistent responding, or serious psychopathology. To select among these various interpretations, elevations should be considered within the context of the non-content-based scales (VRIN-r and TRIN-r) as well as the Infrequent Psychopathological Responses (Fp-r) scale. Scores less than 79T are suggestive of a valid profile with no indication of over-reporting. As noted above, moderate elevations (80 to 119T) may reflect inconsistent responding, genuine psychopathology, emotional distress, or over-reporting. Thus, scores should be interpreted in relation to ones' clinical history as well as VRIN-r, TRIN-r, and Fp-r scale scores (Hoelzle et al., 2012). Across numerous settings, very high elevations (≥120T) are reflective of an invalid profile.

The Fp-r scale comprises 21 items, which are endorsed less than 20 % of the time by the MMPI-2-RF normative and two inpatient psychiatric samples; 14 of the items were retained from the from the MMPI-2 Fp scale (Hoelzle et al., 2012). Fp and Fp-r have been found to be highly correlated. Consequently, the interpretation guidelines for Fp-r have remained the same (Hoelzle et al., 2012). Indeed, Sellbom et al. (2010) recently reaffirmed that previous interpretive guidelines (Ben-Porath & Tellegen, 2008) can still be used successfully to identify over-reporting within a forensic context.

The Infrequent Somatic Responses (Fs) scale (Wygant, Ben-Porath, & Arbisi, 2004) is a 16-item scale that was developed specifically for the MMPI-2-RF. This scale helps identify noncredible reports of somatic symptoms. Wygant et al. (2004) employed a *rare-symptom* approach to identify

items endorsed by less than 25 % of patients in large archival medical and chronic pain samples. Items with content describing somatic and physical complaints were retained from the large list of infrequently endorsed items. Thus, this scale includes items with somatic content rarely endorsed by medical and chronic pain patients. Therefore, elevations on this scale (>80T) reflect the potential for exaggerated somatic complaints. Nevertheless, scores should be interpreted within the context of the individual's medical history. Ben-Porath and Tellegen (2008/2011) suggest that scores between 80 and 99T could either be indicative of genuine medical conditions or exaggerated somatic complaints. Within this range, a thorough review of an individual's medical history should be conducted to aid in determining whether the moderate elevation reflects a genuine medical condition or an exaggerated report of somatic symptoms. The somatic substantive scales of the MMPI-2-RF [somatic complaints (RC1), malaise (MLS), gastrointestinal complaints (GIC), head pain complaints (HPC), neurological complaints (NUC), and cognitive complaints (COG)] should be interpreted with caution at this level. When scores are greater than 100T, the somatic symptoms endorsed by the test taker are rarely described by persons with genuine medical problems. Consequently, their report of somatic complaints on the test should be interpreted with significant caution and likely represents over-reporting. Indeed, in a review of normative, clinical, medical, and personal injury litigant samples, Greene (2011) found that Fs scores above 100T were found in fewer than 2 % of profiles. Moreover, in an examination of somatic malingering, Sellbom, Wygant, and Bagby (2012) found that Fs was the most sensitive, but Fp-r was the most specific.

Similar to the MMPI-2's FBS, the Symptom Validity (FBS-r) scale is designed to assess noncredible somatic and neurocognitive complaints and has retained 30 of the original items (Wygant et al., 2011). However, unlike the original, FBS-r shares 12 items with other validity scales (Hoelzle et al., 2012). Elevated scores on FBS-r are indicative of over-reporting of cognitive and somatic deficits. In particular, scores between 80 and 99T indicate moderate elevation and possible over-reporting. Therefore, the individual's background should be considered. Within this score range, the somatic and cognitive specific problems scales (RC1, MLS, GIC, HPC, NUC, COG) will likely need to be interpreted with caution. High elevations (≥100T) are indicative of exaggerated or unusual reporting of cognitive or somatic symptoms (Ben-Porath & Tellegen, 2008). Scores in this range have been observed in less than 1 % of profiles from normative, clinical, litigating, and medical samples (Greene, 2011). If the overall profile is still considered to be valid, interpretation of the scores on RC1 and the cognitive and somatic specific problems scales will require significant caution (Graham, 2011).

More recently, the Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant, & Green, 2007) was added as an official

32

Psychol. Inj. and Law (2015) 8:27–39

MMPI-2-RF over-reporting scale. It was developed to identify self-reported symptoms (regardless of item content) associated with poor performance on cognitive PVTs. Employing a large sample of evaluees who completed the MMPI-2 as well as several cognitive PVTs, primarily in worker's compensation board evaluations, Gervais et al. (2007) identified 28 items that significantly discriminated between those who failed and passed these measures. Although FBS scores have been found to be associated with PVT performance (e.g., Wygant et al., 2007), RBS has been found to outperform FBS and other MMPI-2-RF validity scales in predicting poor performance on PVTs (Gervais, Ben-Porath, Wygant, & Green, 2007). Of note, in a recent study by McBride et al. (2013), RBS was found (along with cognitive PVTs) not to be significantly affected by the presence of bona fide brain damage. While one would never substitute the RBS scale of the MMPI-2-RF for a cognitive PVT, elevations on this scale in conjunction with poor performance on PVTs would only enhance a conclusion of the presence of response bias.

MMPI-2-RF in Psychological Injury and Related Evaluations

Although the MMPI-2-RF was published only 7 years ago, it has already amassed an impressive amount of empirical investigation. A large proportion of these articles specifically investigate the test's validity scales in psychological injury (and related) types of evaluations. Wygant et al. (2009) examined F-r, Fs, and FBS-r in civil forensic settings, using a known-groups design with cognitive performance validity tests (PVTs) as response bias criteria. They found that these scales discriminated between passing and failing participants, demonstrating adequate sensitivity and notable specificity. More recently, Gervais, Wygant, Sellbom, and Ben-Porath (2011) used PVTs to investigate the utility of the MMPI-2-RF scales in a disability setting. Similarly, PVT failure was associated with significant elevations on the aforementioned over-reporting scales, exhibiting the ability of the scales to detect the over-reporting of emotional, somatic, and neurocognitive complaints.

The utility of the MMPI-2-RF validity scales have also been examined in the prediction of structured malingering criteria in samples of compensation-seeking individuals. Tarescavage, Wygant, Gervais, and Ben-Porath (2013) examined the MMPI-2-RF validity scales in relation to the MND criteria in a sample of non-head injury compensation evaluations. Those authors found that higher scores on MMPI-2-RF validity scales, particularly RBS were associated with probable and definite MND. Moreover, they examined the combined use of the validity scales and found that the overall accuracy in identifying MND improved when multiple scales were employed. In a separate study employing TBI patients, Schroeder et al. (2012) found that the MMPI-2-RF over-reporting scales exhibited excellent

classification accuracy in discriminating TBI litigants classified as probable malingers from non-malingerers. Finally, Wygant et al. (2011) found that Fs and FBS-r were good at identifying noncredible neurocognitive and somatic symptoms in a sample of litigants undergoing compensation-seeking evaluations for disability who were classified with the MND and MPRD criteria. Furthermore, regression-based analyses have supported the incremental validity of the MMPI-2-RF F-r, Fp-r, Fs, and FBS-r in relation to the MMPI-2 Fp and FBS in the prediction of exaggerated memory complaints (Gervais, Ben-Porath, Wygant, & Sellbom, 2010).

The MMPI-2-RF validity scales have also been useful in identifying malingered symptoms of specific disorders. In an examination of the ability of over-reporting scales to distinguish feigned major depressive disorder (MDD), schizophrenia, or PTSD from genuine psychiatric patients, Marion, Sellbom, and Bagby (2011) found that the scales were able to distinguish simulators from actual patients. Specifically, they were able to identify most simulators regardless of the sophistication of their training. Furthermore, in an effort to determine how well specific knowledge could affect the ability of the scales to identify over-reporting of PTSD symptomology, Goodwin et al. (2013) examined the responses of veterans seeking disability compensation and mental health professionals instructed to feign symptoms of PTSD. Even though the mental health professionals were able to feign symptoms in a more sophisticated manner, the MMPI-2-RF validity scales were still effective in distinguishing feigners from actual sufferers. Additionally, Mason et al. (2013) examined the accuracy of the validity scales to detect malingered PTSD symptoms in relation to genuine PTSD and various degrees of random responding. These authors compared undergraduates randomly assigned to either respond honestly, feign PTSD, partially randomly respond, or fully randomly respond on the MMPI-2-RF with veterans diagnosed with PTSD based on a structured clinical interview and malingering assessment. The validity scales were able to correctly classify 80 % of the genuine PTSD patients and 73 % of the subjects feigning PTSD. Their results supported the use of the validity scales in distinguishing feigning from genuine responding and veterans with PTSD. They found weaker support for the MMPI-2-RF validity scales detecting partially random responding.

Personality Assessment Inventory

The Personality Assessment Inventory (PAI; Morey, 1991) is a 344-item self-report measure designed to assess a broad range of personality factors and psychopathological symptoms as well as maladaptive personality traits (Morey, 1991/ 2007). The PAI utilizes a Likert-scale design with four response options. The test comprises 22 scales in total,

organized into validity, clinical (most with 3–4 subscales), treatment consideration, and interpersonal scales. The PAI is appropriate for individuals aged 18 years and older. Its items are understood and applicable across cultures, and it has been translated into several languages.

There has been substantially less empirical work examining the three primary PAI over-reporting indices than the MMPI-2 validity scales. Nevertheless, in their review of the measure, Sellbom and Bagby (2008) noted strong support for the use of the PAI validity indicators in assessing response bias, based on research employing both known-groups and simulation designs (see Bagby, Nicholson, Bacchiochi, Ryder, & Bury, 2002, Baity, Siefert, Chambers, & Blais, 2007; Boccaccini, Murrie, & Duncan, 2006; Edens, Poythress, & Watkins-Clay, 2007; Kucharski, Toomey, Fila, & Duncan, 2007; Liljequist, Kinder, & Shinka, 1998; Morey & Lanier, 1998).

PAI Over-Reporting Indicators

The Infrequency (INF) scale was created to evaluate atypical response patterns resulting from confusion, carelessness, reading difficulties, or random responding. More specifically, it uses an improbable symptom-detection strategy, containing eight extremely unusual items for most test takers. The items are balanced, so that half are considered "false" and the other "very true" for the majority of test takers (Morey, 2007). Although the items are mainly a measure of careless responding, elevations on this scale can also be a consequence of idiosyncratic responding. Furthermore, the items are evenly distributed throughout the PAI, so problematic responding can be detected at any point of the test administration. In contrast to the MMPI-2's F scale, these items were written without using bizarre content. Morey (1991) selected the scale's items on the basis of infrequent endorsement in community and clinical samples.

In his interpretive guidelines, Morey (2007) reported that moderate elevations on this scale (60 to 74T) can be attributed to somewhat atypical responding. In particular, when scores are in this range, clinicians should consider reading difficulties, confusion, scoring errors, idiosyncratic item interpretation, random responding, or failure to follow testing instructions as possible reasons for the elevation. Consequently, the remaining protocol should be interpreted with caution. High elevations (≥75T) indicate that the respondent did not properly attend to the test items. As noted earlier, potential sources of error should be investigated. Nevertheless, highly elevated scores suggest the test results are best assumed to be invalid, and the remainder of the PAI scales should not be interpreted (Morey, 2007).

Another over-reporting scale, Negative Impression Management (NIM), contains nine items. The content of the items reflects extremely bizarre and unlikely symptoms, and the

scale is thought to assess a respondent's tendency to present a negative impression of psychological functioning. Similar to the MMPI-2's F and MMPI-2-RF's F-r scales, it utilizes a *rare-symptoms* approach to identify an exaggerated or overly negative response style. Literature regarding malingering, factitious disorder, and pseudopsychosis guided item development (Morey, 2007). Originally, Morey (1991) investigated its utility and efficacy with an undergraduate sample by instructing participants to feign certain disorders. These profiles were then compared to those in clinical and nonclinical populations under normal conditions. Morey found the NIM cut score of ≥73T to be the most effective in distinguishing between patients and simulators. However, other researchers have used up to 110T as a cutoff (Blanchard, McGrath, Pogge, & Khadivi, 2003; Rogers, Sewell, Morey, & Ustad, 1996). Furthermore, in these initial validation studies, Morey (1991) reported that individuals scoring higher than the critical cutoff were 14.7 times more likely to be in the feigning simulator group rather than a member of the clinical sample. However, Morey (2007) also underscored that although NIM is a strong predictor of feigning, it "is not a malingering scale per se" (p. 29). Like the MMPI-2's F scale, it also can be elevated in the presence of severe disorders. Nevertheless, it can similarly elevate as a result of idiosyncratic responding; however, this would also be substantiated by an elevation on INF.

In the PAI's interpretive guidelines, Morey (2007) reports that NIM scores below 73T indicate that there is little to no distortion in a pathological direction. Moderate elevations (73 to 83T) suggest some exaggeration regarding symptomology or misfortunes. Thus, scores within this range should be interpreted with caution. Elevations ranging from 84 to 91T are still considered moderate but may also reflect a "cry for help." High scores (≥92T) on this scale pose the strong possibility of either careless responding, highly negative self-portrayal, or malingering. As a result, the profile would be considered invalid. In their review, Sellbom and Bagby (2008) report, however, that regardless of the cutoff score, sensitivity can be problematic for this scale.

The Malingering Index (MAL; Morey, 1996) was developed as a means of more directly examining symptom exaggeration per se than NIM. This index assesses profile distortions using eight characteristics, which are distinguishable in the PAI profile and have frequently been found in the profiles of feigners (Morey, 2003; Sellbom & Bagby, 2008; Thomas, Hopwood, Orlando, Weathers, & McDevitt-Murphy, 2012). The MAL utilizes an *unlikely pattern of psychopathology* detection strategy, in which elaborate symptom combinations are examined (Sellbom & Bagby, 2008). This particular strategy uses elaborate and unlikely symptom combinations, which may be common to a clinical population, but rarely occur together (Rogers, 2008). Such configurations are more likely to occur with noncredible reporting. When an unusual

response pattern occurs on more than one of the configural indicators, it is thought to potentially reflect a distorted response style. Factor analysis revealed that MAL assesses reports of unusual psychotic symptoms and a factor characterized by negative attitudes toward oneself and the world (Veltri & Williams, 2013). Morey (2007) recommends a raw score of ≥3 (i.e., 84T) to be an effective cut score as a screen for possible malingering. Scores of ≥5 (i.e., 111T) are highly unusual in clinical populations and usually occur when severe mental disturbances are feigned. However, when milder forms of pathology (e.g., depression, anxiety) are simulated, MAL's sensitivity is decreased. Therefore, when malingering of milder clinical disorders is suspected, a lower cutoff score should be considered (Morey, 2007).

Rogers et al. (1996) developed the Rogers Discriminant Function (RDF) to empirically identify malingering while remaining unrelated to particular psychopathology or deliberate negative impression management (Morey, 2007). Specifically, RDF is comprised of 20 PAI scales as well as subscales thought to best classify feigning. Similar to the MAL index, it adheres to the *unlikely patterns of psychopathology* detection strategy (Sellbom & Bagby, 2008). Morey (1996) recommends that RDF scores above 0 are suggestive of malingering, whereas anything below 0 does not indicate an attempt to distort the profile.

The RDF has demonstrated the most variability of the PAI over-reporting scales as regards to its utility (Hawes & Boccaccini, 2009). Some research has indicated that it can discriminate between individuals feigning psychopathology and patients with genuine disorders (Rogers et al., 1996). Bagby et al. (2002) compared the validity scores of participants instructed to feign a mental disorder, in which half of the sample was coached. The profiles were then compared with those of patients with genuine psychopathology. The RDF exhibited superior detection of noncredible responding over NIM and MAL. Specifically, in a meta-analysis, Hawes and Boccaccini (2009) discovered that although NIM generated the largest overall effect sizes when distinguishing coached and uncoached feigning, MAL and RDF also produced moderate to large effect sizes. Therefore, given that the RDF index is unrelated to psychopathology and still able to identify individuals in feigning simulations, it has been suggested that it may be a reasonably "pure" marker of feigning on the PAI (Thomas et al., 2012).

More recently, the support for the RDF to accurately detect feigned disorders has been less robust. In one simulation study, the RDF failed to significantly differentiate the honest from feigning groups (Boccaccini et al., 2006). In another study examining feigned PTSD, the RDF index performed more poorly than it has in other feigning studies and may not generalize well to PTSD research (Thomas et al., 2012). Kucharski et al. (2007) found that RDF was not useful in identifying malingered psychiatric disorders among actual

criminal defendants, and it was not correlated with the Structured Interview of Reported Symptoms (SIRS; Rogers, 1986; Rogers, Bagby, & Dickens, 1992) total scores. When examining studies employing criterion group designs, the RDF was no better than chance in its detection of over-reporting, whereas NIM and MAL maintained moderate effect.

Previous research has generally supported the use of the PAI validity indicators in identifying an exaggerated response style in forensic settings. When using the SIRS as a criterion measure, NIM was found to perform just as well as the MMPI-2 F scale and was the most effective PAI scale in identifying purposeful exaggeration (Boccaccini et al., 2006). Sellbom and Bagby (2008) similarly concluded that NIM appears to show higher correlations with external measures of feigning (e.g., SIRS), whereas MAL is moderately to strongly correlated with external measures and the RDF is the least correlated with external measures of feigning. In another study, Kucharski et al (2007) found that with a cutoff of 84T, NIM was reasonably accurate in its detection of suspected malingerers (approximately 87 %) in criminal defendants feigning psychiatric disorders. They further found that both the NIM and MAL index scores were significantly correlated with SIRS total scores, whereas RDF was not. Conversely, Bagby et al. (2002) were unable to find a significant difference between NIM scores of undergraduates instructed to feign and psychiatric patients.

PAI in Psychological Injury and Related Evaluations

Although few studies have directly examined the PAI in psychological injury evaluations, research has examined its validity scales in identifying feigned PTSD. Rogers, Gillard, Wooley, and Ross (2012) were also able to identify 72 % of genuine PTSD respondents correctly. However, they used a cutoff of 84T. Furthermore, at the expense of sensitivities, they were able to achieve very high specificities (>0.95) on the NIM, MAL, and RDF indices. Thomas et al. (2012) found that with a score greater than 69T, NIM was able to correctly classify 75 % of a sample comprising PTSD patients, community members, and university undergraduates. Consistent with previous research (Lange, Sullivan, & Scott, 2010), at a cutoff score of 2, MAL was highly sensitive (0.94), but its specificity was low (0.52). In their meta-analysis of the PAI validity scales, Hawes and Boccaccini (2009) reported that a cut score ≥3 on MAL provided the highest overall classification rate (0.71) but had a specificity of 0.86 and sensitivity of 0.58. Therefore, they concluded that contrary to Morey's (2007) interpretive guidelines, a MAL score of ≥4 was the most likely indicator of feigning.

Keiski, Shore, Hamilton, and Malec (in press) utilized an analogue simulation design to compare patients with traumatic brain injuries (TBIs) to those feigning either specific cognitive and somatic symptoms related to TBI or a wide array of

related cognitive, somatic, and psychiatric symptoms. NIM, MAL, and RDF were able to distinguish between the simulation and TBI groups. While these three indicators were sensitive to simulated TBI symptoms, they were less sensitive in detecting feigned specific somatic and cognitive TBI symptoms than broad somatic, cognitive, and emotional symptoms associated with TBI. Furthermore, the over-reporting scales were somewhat sensitive to TBI simulation and generated large effect sizes. Simulators produced the highest mean validity scores on NIM. Their findings are consistent with previous research in which high scores on NIM have been associated with poor effort on PVTs or compensation-seeking populations (see Lange, Pancholi, Bhagwat, Anderson-Barnes, & French, 2012). For instance, higher NIM scores were observed for compensation-seeking patients with mild TBIs than for patients who were not compensation seeking but also had mild TBIs (Whiteside, Galbreath, Brown, & Turnbull, 2012). Furthermore, in an examination of the PAI validity scales in relation to cognitive effort as measured by the Test of Memory Malingering (TOMM; Tombaugh, 1996), the INF and NIM scales were found to be significantly related to the TOMM (Whiteside, Dunbar-Mayer, & Waters, 2009). However, INF was only significantly related to the TOMM trial 1, whereas NIM correlated significantly with the first, second, and retention trials. This suggests that poor effort on the TOMM may be associated with self-report response bias on the PAI. Similarly, Lange et al. (2012) employed a criterion-groups design based on cognitive PVTs and found that patients who passed PVTs obtained lower NIM scores, regardless of mild or severe TBI classification, than those failing PVTs with mild TBIs.

Comparing the MMPI-2 and PAI

There have been numerous studies comparing the efficacy of the MMPI-2 and PAI in the detection of feigned disorders. In particular, Lange et al. (2010) examined both measures' ability to detect feigned depression and PTSD. They found that although all the MMPI-2 and PAI validity indicators exhibited high specificity, PPP, and NPP values, there were differences between the measures with regard to sensitivity and concluded that the MMPI-2 was superior in its ability to detect feigned responses in comparison with the PAI. Eakin, Weathers, Benson, Anderson, and Funderburk (2006) found that both the PAI and MMPI-2 were able to distinguish PTSD from controls, but the MMPI-2 outperformed the PAI in the detection of those instructed to feign the disorder. In particular, the MMPI-2F and Fp scales had higher effect sizes in detecting simulated PTSD than the NIM, MAL, or RDF indicators. Nevertheless, although the MMPI-2 outperformed the PAI in their study, a considerable proportion of those feigning were still able to avoid detection on both tests. The researchers opined that with some coaching and a modest incentive, individuals are better able to successfully feign PTSD than a general mental disorder. They indicate that when feigning instructions in analogue studies are vague and ambiguous, feigning participants are more likely to intensify their over-reporting, thus increasing detection with validity scales. Veltri and Williams (2013) found that, consistent with previous research, participants who had been coached to feign PTSD and generalized anxiety disorder (GAD) were more likely to avoid detection than those feigning schizophrenia on both the PAI and MMPI-2.

One possible explanation for the apparent superiority of the MMPI-2 over the PAI in detecting feigned PTSD has to do with the construction of the measures. The MMPI was developed using an empirical criterion-keyed method; therefore, many of the items are not face valid, which can make it difficult to discern which items belong to scales assessing PTSD symptoms. In contrast, the PAI was developed by employing a construct validation approach, so many of its items are face valid (Eakin et al., 2006).

General Conclusions Concerning the Assessment of Response Bias in Psychological Injury Evaluations

Forensic psychological injury evaluations require extensive assessment and consideration of malingering and response bias. Psychological injury evaluations are complicated in that the individual may report symptoms across multiple domains of functioning (e.g., psychological/psychiatric, somatic, neurocognitive). All three tests discussed in this paper have shown their clinical utility in forensic psychological evaluations and can be incorporated into an assessment battery with confidence. However, whereas all three measures are able to capture the psychological symptoms and personality traits relevant to psychological injury evaluations, the MMPI-2 and MMPI-2-RF may be better suited in the assessment of response bias. As Young (2014) has noted, the number of studies investigating the use of the PAI in psychological injury settings is much smaller than those concerning the MMPI (Young, 2014). While the number of studies investigating a particular test should not be the only indicator of strength in using that measure in a forensic evaluation, it does potentially increase one's confidence in using the measure in a forensic setting, particularly if expert testimony will be required. Due to the heterogeneous symptoms often presented during these evaluations, it is important to capture elements of feigning across all areas of functioning. The MMPI-2 and MMPI-2-RF have been more thoroughly examined in relation to the criteria for MND and MPRD (Bianchini et al., 2008; Greve et al., 2006; Schroeder et al., 2012; Tarescavage et al., 2013; Wygant et al., 2011). In comparison, the PAI has much fewer studies examining the relation between its validity indicators with cognitive PVTs, MND, and MPRD. To our knowledge, only one study has examined the PAI in relation to MPRD and none have examined the test in

relation to the MND. In a simulation design examining malingered pain-related disability, Hopwood, Orlando, and Clark (2010) found that although NIM, MAL, and RDF demonstrated significant effects for distinguishing between self-reported pain and malingered pain-related disability, the scales' detection ability were not sufficiently sensitive; thus, they were not recommended for routine clinical use.

Few studies have directly compared the MMPI-2 and MMPI-2-RF validity scales. While Gervais et al. (2010) found that the MMPI-2-RF validity scales (particularly RBS) added incrementally to the MMPI-2 validity scales in predicting exaggerated memory complaints, additional studies are needed before any definitive conclusions can be drawn about the comparative merits of each measures' validity scales. Hoelzle et al. (2012) point out that in many ways, the validity scales from the MMPI-2 and MMPI-2-RF complement one another. Conversely, Tellegen and Ben-Porath (2008/2011) show that the correlations between the MMPI-2 and MMPI-2-RF validity scales are quite high, suggesting that these scales work in a similar fashion.

Several factors must be considered when deciding whether to employ either the MMPI-2 or MMPI-2-RF in a psychological injury evaluation. The MMPI-2 has been in use for a much longer period of time than the MMPI-2-RF, which may be associated with more familiarity for both the clinician and the court where MMPI testimony is offered on occasion. That would be sufficient for some to decide to use the MMPI-2 over the MMPI-2-RF. However, while the MMPI-2-RF has only been published since 2008, it has already amassed an impressive amount of empirical research and would likely have no problems withstanding a challenge to its admission in expert testimony (see Ben-Porath (2012) and Sellbom (2012)).

In comparing the MMPI-2 and MMPI-2-RF in their ability to assess response bias in psychological injury evaluations, these authors would recommend the MMPI-2-RF. The reasoning behind this recommendation is twofold and stems from primarily practical considerations. First, the MMPI-2-RF is substantially shorter in length than the MMPI-2 without sacrificing much in terms of clinical coverage. This can be important in a forensic evaluation where time is valuable. For instance, a clinician could employ the MMPI-2-RF in an evaluation in conjunction with a trauma-specific inventory like the TSI-2, if it is relevant to the case (e.g., PTSD evaluation), and still administer fewer items than the MMPI-2. Second, the standard validity scales of the MMPI-2-RF appear to provide better overall coverage of symptom exaggeration, although again, this issue needs to be fleshed out with additional research. Conceptually speaking, while both versions of the test include similar measures of F/F-r, Fp/Fp-r, and FBS/FBS-r, the two additional and unique over-reporting validity scales on the MMPI-2-RF offer a unique examination of exaggerated somatic symptoms (with Fs), utilizing a *rare-symptoms* approach, and symptoms empirically associated with poor performance on cognitive response bias indicators (with RBS). As noted earlier, these two scales

have been found to provide incremental validity in the assessment of feigned somatic complaints and cognitive dysfunction. In some cases, it may also be useful to employ both the MMPI-2-RF and PAI in a forensic psychological evaluation. The clinical constructs across these two measures are distinct enough that they are likely to complement one another. Moreover, since each measure is significantly shorter than the MMPI-2, both can be utilized in an evaluation and the total number of items administered to the client would only be 115 more than the MMPI-2 by itself.

Self-report measures like those discussed in this article are routinely utilized in forensic psychological injury evaluations, both for their economy of use and their broad coverage of psychopathological symptoms and personality traits. The aforementioned measures all have shown significant empirical support and furthermore are uniquely suited to be incorporated into these types of evaluations. Nevertheless, a thorough understanding of the working and interpretation of each scale is necessary to be able to accurately integrate these measures into expert witness testimony that will withstand rigorous cross examination.

References

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: the Infrequency-Psychopathology Scale, F(p). *Psychological Assessment, 7,* 424–431.

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2006). The ability of the MMPI-2 to detect feigned PTSD within the context of compensation seeking. *Psychological Services, 3,* 249–261.

Bagby, R., Nicholson, R. A., Bacchiochi, J. R., Ryder, A. G., & Bury, A. S. (2002). The predictive capacity of the MMPI-2 and PAI validity scales and indexes to detect coached and uncoached feigning. *Journal of Personality Assessment, 78,* 69–86.

Bagby, R., 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.

Baity, M. R., Siefert, C. J., Chambers, A., & Blais, M. A. (2007). Deceptiveness on the PAI: a study of naive faking with psychiatric inpatients. *Journal of Personality Assessment, 88,* 16–24.

Ben-Porath, Y. S. (2012). Addressing challenges to MMPI-2-RF-based testimony: questions and answers. *Archives of Clinical Neuropsychology, 27,* 691–705.

Ben-Porath, Y. S., Graham, J. R., & Tellegen, A. (2009). *The MMPI-2 symptom validity scale (FBS): development, research findings, and interpretive recommendations.* Minneapolis: University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen A. (2008/2011). *MMPI-2-RF: User's guide for reports* (2nd ed.). Minneapolis: University of Minnesota Press. *MMPI-2-RF: Manual for administration and scoring* (rev. ed.) Minneapolis: University of Minnesota Press.

Bianchini, K. J., Etherton, J. L., Greve, K. W., Heinly, M. T., & Meyers, J. E. (2008). Classification accuracy of the MMPI-2 in the detection of pain-related malingering. *Assessment, 15,* 435–449.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). Review article: On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *The Spine Journal, 5,* 404–417.

Blais, M. A., Baity, M. R., & Hopwood, C. J. (2010). *Clinical applications of the Personality Assessment Inventory.* New York: Routledge/Taylor & Francis Group.

Blanchard, D. D., McGrath, R. E., Pogge, D. L., & Khadivi, A. (2003). A comparison of the PAI and MMPI-2 as predictors of faking bad in college students. *Journal of Personality Assessment, 80,* 197–205.

Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal-forensic sample with the Personality Assessment Inventory. *Psychological Assessment, 18,* 415–423.

Briere, J. (1995). *Trauma Symptom Inventory professional manual.* Odessa, FL: Psychological Assessment Resources.

Briere, J. (2010). *Trauma Symptom Inventory, 2nd ed. (TSI-2) professional manual.* Odessa, FL: Psychological Assessment Resources.

Burchett, D. L., & Bagby, R. M. (2014). Multimethod assessment of response distortion: Integrating data from interviews, collateral records, and standardized assessment tools. In C. J. Hopwood & R. F. Bornstein (Eds.), *Multimethod clinical assessment* (pp. 345–378). New York: Guilford Press.

Burchett, D. L., & Ben-Porath, Y. S. (2010). The impact of overreporting on MMPI-2-RF substantive scale score validity. *Assessment, 17,* 497–516.

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law, 7,* 197–205.

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. (2010). Personality assessment from the nineteenth to the early twenty-first century: past achievements and contemporary challenges. *Annual Review of Clinical Psychology, 6,* 1–20.

Butcher, J. N., Arbisi, P. A., Atlis, M. M., & McNulty, J. L. (2003). The construct validity of the Lees-Haley Fake Bad Scale: does this scale measure somatic malingering and feigned emotional distress? *Archives of Clinical Neuropsychology, 18,* 473–485.

Butcher, J. N., Gass, C. S., Cumella, E., Kally, Z., & Williams, C. L. (2008). Potential for bias in MMPI-2 assessments using the Fake Bad Scale (FBS). *Psychological Injury and Law, 1,* 191–209.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2): manual for administration, scoring, and interpretation, revised edition.* Minneapolis: University of Minnesota Press.

Butcher, J. N., & Miller, K. B. (1999). Personality assessment in personal injury litigation. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (2nd ed., pp. 104–126). New York: Wiley.

Crawford, E. F., Greene, R. L., Dupart, T. M., Bongar, B., & Childs, H. (2006). MMPI-2 assessment of malingered emotional distress related to workplace injury: a mixed group validation. *Journal of Personality Assessment, 86,* 217–221.

Dean, A. C., Boone, K. B., Kim, M. S., Curiel, A. R., Martin, D. J., Victor, T. L., … Lang, Y. K. (2008). Examination of the impact of ethnicity on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Fake Bad Scale. *The Clinical Neuropsychologist, 22,* 1054–1060.

Demakis, G. J., & Elhai, J. D. (2011). Neuropsychological and psychological aspects of malingered posttraumatic stress disorder. *Psychological Injury and Law, 4,* 24–31.

Eakin, D. E., Weathers, F. W., Benson, T. B., Anderson, C. F., & Funderburk, B. (2006). Detection of feigned posttraumatic stress disorder: a comparison of the MMPI-2 and PAI. *Journal of Psychopathology and Behavioral Assessment, 28,* 145–155.

Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: a comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment, 88,* 33–42.

Efendov, A. A., Sellbom, M., & Bagby, R. (2008). The utility and comparative incremental validity of the MMPI-2 and Trauma symptom inventory validity scales in the detection of feigned PTSD. *Psychological Assessment, 20,* 317–326.

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,* 541–559.

Gass, C. S., Williams, C. L., Cumella, E., Butcher, J. N., & Kally, Z. (2010). Ambiguous measures of unknown constructs: the MMPI-2 Fake Bad Scale (aka Symptom Validity Scale, FBS, FBS-r). *Psychological Injury and Law, 3,* 81–85.

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., & 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., Wygant, D. B., Sellbom, M., & Ben-Porath, Y. S. (2011). Associations between symptom validity test failure and scores on the MMPI-2-RF validity and substantive scales. *Journal of Personality Assessment, 93,* 508–517.

Goodwin, B. E., Sellbom, M., & Arbisi, P. A. (2013). Post-traumatic stress disorder in veterans: the utility of the MMPI-2-RF Validity Scales in detecting over-reported symptoms. *Psychological Assessment, 25,* 671–678.

Graham, J. R. (2011). *The MMPI-2: assessing personality and psychopathology* (5th ed.). New York: Oxford University Press.

Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: an interpretive manual* (3rd ed.). New York: Allyn Bacon.

Greiffenstein, M. F. (2010). The MMPI-2 Symptom Validity Scale (FBS) not influenced by medical impairment: a large sleep center investigation. *Assessment, 17,* 269–277.

Greve, K. W., Bianchini, K. J., Love, J. M., Brennan, A., & Heinly, M. T. (2006). Sensitivity and specificity of MMPI-2 validity scale and indicators to malingered neurocognitive dysfunction in traumatic brain injury. *The Clinical Neuropsychologist, 20,* 491–512.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory manual.* New York: Psychological Corporation.

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: a meta-analytic review. *Psychological Assessment, 21,* 112–124.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Participants, C. (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., & Stanczak, S. R. (2008). Comparison of the Lees-Haley Fake Bad Scale, Henry-Heilbronner Index, and Restructured Clinical Scale 1 in identifying noncredible symptom reporting. *The Clinical Neuropsychologist, 22,* 919–929.

Hoelzle, J. B., Nelson, N. W., & Arbisi, P. A. (2012). MMPI-2 and MMPI-2-Restructured Form validity scales: complementary approaches to evaluate response validity. *Psychological Injury and Law, 5,* 174–191.

Hopwood, C. J., Orlando, M. J., & Clark, T. S. (2010). The detection of malingered pain-related disability with the Personality Assessment Inventory. *Rehabilitation Psychology, 55,* 307–310.

Keiski, M. A., Shore, D. L., Hamilton, J. M., & Malec, J. F. (in press). Simulations of traumatic brain injury symptoms on the Personality Assessment Inventory: An analogue study. *Assessment*.

Kucharski, L., Toomey, J. P., Fila, K., & Duncan, S. (2007). Detection of malingering of psychiatric disorder with the Personality Assessment Inventory: an investigation of criminal defendants. *Journal of Personality Assessment, 88,* 25–32.

Lange, R. T., Pancholi, S., Bhagwat, A., Anderson-Barnes, V., & French, L. M. (2012). Influence of poor effort on neuropsychological test performance in U. S. Military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 34,* 453–466.

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. (2012). Dialogue: performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 1–7.

Lee, T. C., Graham, J. R., Sellbom, M., & Gervais, R. O. (2012). Examining the potential for gender bias in the prediction of symptom validity test failure by MMPI-2 symptom validity scale scores. *Psychological Assessment, 24,* 618–627.

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.

Liljequist, L., Kinder, B. N., & Shinka, J. A. (1998). An investigation of malingering posttraumatic stress disorder on the Personality Assessment Inventory. *Journal of Personality Assessment, 7,* 322–336.

Marion, B. E., Sellbom, M., & Bagby, R. M. (2011). The detection of feigned psychiatric disorders using the MMPI-2-RF overreporting validity scales: an analog investigation. *Psychological Injury and Law, 4,* 1–12.

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.

Mason, L. H., Shandera-Ochsner, A. L., Williamson, K. D., Harp, J. P., Edmundson, M., Berry, D. R., & High, W. J. (2013). Accuracy of MMPI-2-RF validity scales for identifying feigned PTSD symptoms, random responding, and genuine PTSD. *Journal of Personality Assessment, 95,* 585–593.

McBride, W. F., Crighton, A. H., Wygant, D. B., & Granacher, R. P. (2013). It's not all in your head (or at least your brain): association of traumatic brain lesion presence and location with performance on measures of response bias and forensic evaluation. *Behavioral Sciences and the Law, 31,* 774–778.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136,* 450–470.

Miller, L., Sadoff, R. L., & Dattilio, F. M. (2011). Personal injury: the independent medical examination in psychology and psychiatry. In E. R. Drogin, F. M. Dattilio, R. L. Sadoff, & T. G. Gutheil (Eds.), *Handbook of forensic assessment: Psychological and psychiatric perspectives* (pp. 277–302). Hoboken, NJ: Wiley.

Morey, L. C. (1991). *The Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (2003). *Essentials of PAI Assessment*. New York: Wiley.

Morey, L. C. (2007). *The Personality Assessment Inventory professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources.

Morey, L. C., & Lanier, W. V. (1998). Operating characteristics of six response distortion indicators for the Personality Assessment Inventory. *Assessment, 5,* 203–214.

Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A., & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 Symptom Validity Scale (FBS): verified utility in forensic practice. *The Clinical Neuropsychologist, 24,* 701–724.

Nelson, N. W., Sweet, J. J., & Demakis, G. J. (2006). Meta-analysis of the MMPI-2 Fake Bad Scale: utility in forensic practice. *The Clinical Neuropsychologist, 20,* 39–58.

Nichols, D. S. (2011). *Essentials of MMPI-2 assessment* (2nd ed.). New York: Wiley.

Peck, C. P., Schroeder, R. W., Heinrichs, R. J., VonDran, E. J., Brockman, C. J., Webster, B. K., & Baade, L. E. (2013). Differences in MMPI-2 FBS and RBS scores in brain injury, probable malingering, and conversion disorder groups: a preliminary study. *The Clinical Neuropsychologist, 27,* 693–707.

Rogers, R. (1986). *Structured Interview of Reported Symptoms (SIRS)*. Unpublished scale, Clark Institute of Psychiatry, Toronto.

Rogers, R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 14–35). New York, NY, US: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured interview of reported symptoms: professional manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. D. (2012). Evaluation of malingering and related response styles. In R. K. Otto (Ed.), *Comprehensive handbook of psychology: Forensic psychology* (2nd ed., Vol. 11, pp. 517–539). New York: Wiley.

Rogers, R., Gillard, N. D., Wooley, C. N., & Ross, C. A. (2012). The detection of feigned disabilities: the effectiveness of the Personality Assessment Inventory in a traumatized inpatient sample. *Assessment, 19,* 77–88.

Rogers, R., & Granacher, R. P. (2011). Conceptualization and assessment of malingering. In E. R. Drogin, F. M. Dattilio, R. L. Sadoff, & T. G. Gutheil (Eds.), *Handbook of forensic assessment: Psychological and psychiatric perspectives* (pp. 659–678). Hoboken, NJ: Wiley.

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: a discriminant analysis. *Journal of Personality Assessment, 67,* 629–640.

Schroeder, R. W., Baade, L. E., Peck, C. P., VonDran, E., Brockman, C. J., Webster, B. K., & Heinrichs, R. J. (2012). Validation of the MMPI-2-RF validity scales in criterion group neuropsychological samples. *The Clinical Neuropsychologist, 26,* 129–146.

Sellbom, M. (2012). The MMPI-2-RF is ready for the Daubert challenge: evidence, implications, and recommendations for use in court testimony. *Journal of Psychological Practice, 17,* 151–179.

Sellbom, M., & Anderson, J. (2013). The Minnesota multiphasic personality inventory-2. In R. P. Archer & E. M. A. Wheeler (Eds.), *Forensic uses of Clinical Assessment Instruments* (2nd ed., pp. 21–62). New York, NY: Routledge.

Sellbom, M., & Bagby, R. M. (2008). Response styles on multiscale inventories. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (3rd ed., pp. 182–206). New York: Guilford Press.

Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2-RF (restructured form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment, 22,* 22–31.

Sellbom, M., Wygant, D. B., & Bagby, R. M. (2012). Utility of the MMPI-2-RF in detecting non-credible somatic complaints. *Psychiatry Research, 197,* 295–301.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Forum: Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545–561.

Steffan, J. S., Morgan, R. D., Lee, J., & Sellbom, M. (2010). A comparative analysis of MMPI-2 malingering detection models among inmates. *Assessment, 17,* 185–196.

Tarescavage, A., Wygant, D. B., Gervais, R. O., & Ben-Porath, Y. S. (2013). Association between the MMPI-2 Restructured Form (MMPI-2-RF) and malingered neurocognitive dysfunction among

Psychol. Inj. and Law (2015) 8:27–39

39

non-head injury disability claimants. *The Clinical Neuropsychologist, 27,* 313–335.

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

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 Restructured Clinical (RC) Scales: development, validation, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Thomas, K. M., Hopwood, C. J., Orlando, M. J., Weathers, F. W., & McDevitt-Murphy, M. E. (2012). Detecting feigned PTSD using the Personality Assessment Inventory. *Psychological Injury and Law, 5,* 192–201.

Thomas, M. L., & Youngjohn, J. R. (2009). Let's not get hysterical: Comparing the MMPI-2 validity, clinical, and RC scales in TBI litigants tested for effort. *The Clinical Neuropsychologist, 23,* 1067–1084.

Tolin, D. F., Steenkamp, M. M., Marx, B. P., & Litz, B. T. (2010). Detecting symptom exaggeration in combat veterans using the MMPI–2 symptom validity scales: a mixed group validation. *Psychological Assessment, 22,* 729–736.

Tombaugh, T. N. (1996). *Test of memory malingering (TOMM)*. New York: Multi Health Systems.

Veltri, C. O. C., & Williams, J. E. (2013). Does the disorder matter? Investigating a moderating effect on coached noncredible overreporting using the MMPI-2 and PAI. *Assessment, 20,* 199–209.

Whiteside, D. M., Dunbar-Mayer, P., & Waters, D. P. (2009). Relationship between TOMM performance and PAI validity scales in a mixed clinical sample. *The Clinical Neuropsychologist, 23,* 523–533.

Whiteside, D., Galbreath, J., Brown, M., & Turnbull, J. (2012). Differential response patterns on the personality assessment inventory (PAI) in compensation-seeking and non-compensation-seeking mild traumatic brain injury patients. *Journal of Clinical and Experimental Neuropsychology, 34,* 172–182.

Wiggins, C. W., Wygant, D. B., Hoelzle, J., & Gervais, R. O. (2012). The more you say, the less it means: over-reporting and attenuated criterion validity in a forensic disability sample. *Psychological Injury and Law, 5,* 162–173.

Williams, C. L., Butcher, J. N., Gass, C. S., Cumella, E., & Kally, Z. (2009). Inaccuracies about the MMPI-2 Fake Bad Scale in the reply by Ben-Porath, Greve, Bianchini, and Kaufman (2009). *Psychological Injury and Law, 2,* 182–197.

Wygant, D. B., Anderson, J. L., Sellbom, M., Rapier, J. L., Allgeier, L. M., & Granacher, R. P. (2011). Association of the MMPI-2 Restructured Form (MMPI-2-RF) validity scales with structured malingering criteria. *Psychological Injury and Law, 4,* 13–23.

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., 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., 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.

Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: implications for practice and court*. New York: Springer.