

# Potential for Bias in MMPI-2 Assessments Using the Fake Bad Scale (FBS)

James N. Butcher · Carlton S. Gass · Edward Cumella ·  
Zina Kally · Carolyn L. Williams

Received: 17 November 2007 / Accepted: 17 December 2007  
© Springer Science + Business Media, LLC 2008

**Abstract** The Fake Bad Scale (FBS), developed to identify malingering of emotional distress among claimants seeking compensation for personal injuries, was recently added to the MMPI-2 scoring materials, resulting in its widespread dissemination across the various clinical settings that use the MMPI-2 in psychological evaluations. We examine: (1) questions of item bias in the FBS; (2) how malingering and nonmalingering groups are identified in FBS studies, including whether the research has been broadly inclusive and fully represents the populations assessed by the MMPI-2; (3) the reliability and validity of the FBS; (4) the highly variable recommendations for raw score cut-offs and how they relate to *T* scores; (5) two inpatient groups [men in a tertiary care Veterans Affairs Healthcare System (VA) unit and women in an eating disorders program] who may be inappropriately labeled as malingering by the FBS; (6) the publisher's statement on use of the FBS; and (7) a *Frye*

hearing in Florida where the FBS was excluded from expert testimony, one of three so far. We raise questions about its potential bias against people with disabilities and physical illnesses, women, individuals exposed to highly traumatic situations, and those motivated to present themselves in a favorable light. Psychologists using the FBS for making decisions about clients' motivations need to be aware of the serious problems with the scale's use and the cases of its inadmissibility in court.

**Keywords** Bias · MMPI-2 · Fake Bad Scale · FBS · Symptom Validity · Malingering

## Introduction

The Fake Bad Scale (FBS) was developed using the Minnesota Multiphasic Personality Inventory (MMPI; Butcher et al. 1989) items to detect malingerers in personal injury cases (Lees-Haley et al. 1991). According to the scales' developers, "In the personal injury context, malingering is pretending to be more distressed, more impaired, or more disabled than one is" (p. 203). The FBS is based on its authors' premise that a substantial number of clinicians and plaintiffs' attorneys coach claimants in advance of independent medical evaluations to create false claims. As the first step in developing the FBS, its first author examined the MMPI item pool, rationally selected 43 items to include on the FBS, and then rationally determined the items' scored direction "using unpublished frequency counts of malingerers' MMPI responses and observations of personal injury malingerers" (Lees-Haley et al. 1991, p. 204). Next, FBS scores from five comparison samples were examined: personal injury claimants who "appeared notably credible" ( $n=20$ ); personal injury claimants "who appeared clearly to be malingering" ( $n=25$ ); and

---

J. N. Butcher (✉) · C. L. Williams  
University of Minnesota,  
Minneapolis, MN, USA  
e-mail: butch001@umn.edu

C. L. Williams  
e-mail: willi001@umn.edu

C. S. Gass  
Miami Veterans Affairs Healthcare System,  
Miami, FL, USA  
e-mail: carlton.gass@va.gov

E. Cumella · Z. Kally  
Remuda Programs for Eating Disorders,  
Wickenburg, AZ, USA

E. Cumella  
e-mail: edward.cumella@reumdaranch.com

Z. Kally  
e-mail: zina.kally@reumdaranch.com

medical patients asked to simulate emotional distress within personal injury claims caused by a car crash ( $n=16$ ), toxic substance exposure ( $n=15$ ), or work stress ( $n=36$ ).

This resulted in the selection of the score of 20 on the FBS as the optimal cut-off raw score for detecting malingering in personal injury litigants (Lees-Haley et al. 1991). Finally, a weighting system, rather than the subjects' actual responses to the items, was used to estimate mean FBS raw scores for the normative sample and the 540 psychiatric inpatients included in the MMPI-2 manual (Butcher et al. 1989). This weighting procedure for estimating subject responding to FBS MMPI-2 items led Lees-Haley and colleagues (1991) to estimate that the mean raw scores for normative women would be 13.8 and that for normative men would be 11.7, which, they pointed out, were raw scores below their cut-off of 20. In contrast, their estimated mean raw score for psychiatric inpatients was 19.1 for women and 16.9 for men, which led to a "cautionary note" about the relatively high FBS scores obtained from psychiatric inpatients.

Numerous flaws are evident in the original methodology and conclusions (Lees-Haley et al. 1991) underlying the development of the FBS. These include: (1) inadequate description of the item selection procedures (i.e., use of unavailable and poorly specified frequency data of putative malingerers and no elaboration about the nature of the observations of personal injury malingerers that purportedly aided the selection process); (2) lack of independent and empirical verification of the rationally selected items; (3) no description of the criteria used to determine personal injury claimants who appeared "notably credible"; (4) no description of any sound procedure used to determine those "who appeared clearly to be malingering"; (5) no information about the population from which the subjects were selected; (6) use of small sample sizes that can lead to unstable findings; (7) lack of descriptive information, other than mean age and gender, of the samples; (8) failure to consider gender differences, or other key differences (e.g., disability status, health or mental health problems) in responding to the selected items; and (9) use of weighted estimates to calculate the mean scores of the normative sample and psychiatric inpatients that did not allow presentation of SDs.

These methodological flaws likely contributed to the vagueness of the underlying psychological attributes assessed by the FBS. Although initially developed to assess malingering of emotional distress in personal injury claimants, use of the FBS spread to the assessment of malingering for other types of personal injuries seen in forensic settings, for example, posttraumatic stress disorder (PTSD) and somatic complaints, including the feigning of cognitive deficits during neuropsychological evaluations (e.g., Ben-Porath and Tellegen 2007a, b; Greiffenstein et al. 2004; Larrabee 1998; Lees-Haley 1992; Pearson Assessments 2007; Wygant et al. 2007).

A survey by Sharland and Gfeller (2007) suggested that the FBS was among the top five most frequently used measures of effort or response bias used by doctoral level members of the National Academy of Neuropsychology (NAN), although their small sample ( $n=188$ ) and low response rate of 26% precludes generalizability. Furthermore, their conclusions that 75% of those responding used the FBS are offset by the data presented in their tables (i.e., only 43.2% actually reported using the FBS "often" or "always"). To reach a 75% use rate, respondents who reported "rarely" using the FBS had to be included.

Interestingly, this same sample estimated that in the past year, only 1% of their patients "definitely were malingering", and only 3% "probably were malingering." When canvassed without the qualifiers of their personal experience with patients and the limited time reference of 1 year, when these same respondents were asked the nonspecific question to "estimate the percentage of examinees, in general, who deliberately exaggerate their deficits or feign cognitive impairment in cases involving civil litigation or compensation," their estimate jumped to 20% (Sharland and Gfeller 2007, p. 216). Unfortunately, the respondents were not asked to estimate malingering rates based on their own professional experience with litigants within a limited time frame. However, the estimates of malingering from this survey (1% to 20%) are markedly lower than Lees-Haley et al.'s (1991) indication that 56% of their development sample for the FBS (i.e., 25 of Lees-Haley's 45 patients) was malingering.

Until very recently, the decision to include the FBS in a given patient's MMPI-2 evaluation was left to the discretion of the examining psychologist. However, in January 2007, the test distributor, Pearson Assessments, under the direction of the MMPI-2 publisher, University of Minnesota Press, announced that it would include the FBS scale score as a validity score for every patient's MMPI-2 profile scored using their standard scoring materials. In a press release announcing this decision, Pearson Assessments (2007, p. 1) quoted Lees-Haley, "The FBS catches the more subtle forms of non-credible reporting. For example, the FBS can be helpful in cases where someone with a mild or nonexistent brain injury is trying to appear seriously dysfunctional or disabled but not psychotic." Pearson Assessments (2007, p. 1) also quoted Larrabee, who indicated that the FBS is particularly useful in "neuropsychological situations that have a forensic component—it tends to be sensitive to the exaggeration of injury or illness." The publisher recommended a cut-off raw score above 28 for both men and women, based on current research (Pearson Assessments 2007). However, this research was not reviewed in the most current MMPI-2 manual (Butcher et al. 2001), and we return to this issue below. Pope et al. (2006) recommended caution with respect to using controversial measures in forensic assess-

ment. They cited a number of pitfalls that psychologists encounter in forensic evaluations, including the uncritical acceptance of new psychological measures from a test publisher without examining whether there is adequate research validating its use with the populations and issues at hand.

We will examine whether there is sufficient research supporting the recommended widespread clinical use of the FBS, with a focus on its impact on people with disabilities, including those resulting from brain injuries, mental health problems like PTSD, and those requiring inpatient hospitalization. The FBS was added to the MMPI-2 explicitly to identify people with false personal injury claims, thereby preventing them from receiving financial compensation and/or recovery of medical costs, putting it in the class of psychological tests called high stakes measures (Geisinger 2005). Using as a starting point Geisinger's (2005) description of the traditional wisdom for developing fair and unbiased high stakes psychological measures, we will explore the following:

- 1) Potential item bias in the FBS;
- 2) Definition of the criterion groups used to develop the FBS, including how inclusive and representative they are of the populations administered the MMPI-2 and whether the criteria used to classify people as malingerers are replicable and objective;
- 3) Information about the reliability and validity of the FBS;
- 4) Recommendations regarding raw score cut-offs and how they relate to *T* scores;
- 5) Illustrations of a potential for bias when using the FBS with psychiatric inpatients;
- 6) The publisher's statement on the use of the FBS in lieu of a traditional test manual or interpretive guide;
- 7) Recent *Frye* hearings on the use of the FBS as part of expert testimony, highlighting one of three Florida court cases.

### Item Bias in the FBS

Butcher et al. (2003) rationally grouped the 43 FBS items into five categories: denial of deviant attitudes (15 items), somatic symptoms (14 items), low energy/anhedonia (8 items), tension or stress (4 items), and sleep disturbance (2 items). Many of the items are symptoms of physical health problems (e.g., six of the somatic items are symptoms of gastrointestinal disorders; the low energy items frequently accompany physical illnesses and/or depression) and mental health problems (e.g., FBS item content includes suicidal ideation, and frequent nightmares and concentration problems, often associated with PTSD). A significant number of FBS items overlap with the MMPI-2 clinical

scales Hypochondriasis (13 item overlap with Hs, also referred to as scale 1) and Hysteria (14-item overlap with Hy, also referred to as Scale 3) and the content scale Health Concerns (14-item overlap with HEA), well-validated measures of health concerns or physical symptoms. Greiffenstein et al. (2007) cited an unpublished factor analytic study of FBS items by Fox (2005) in which five factors were found. It is not known if these five factors constitute similar item clusters to those in Butcher et al. (2003).

Independent of alleged faking and related motivations, many factors can influence a person's responses to somatic items, including the effects of a wide variety of physical and psychological disorders. Bona fide physical problems lead to high scores on FBS (Iverson et al. 2002; Meyers et al. 2002). Damage to the central nervous system increases scores on FBS (Greiffenstein et al. 2002) and can lead to raw score elevations that exceed 30 (Greve et al. 2006). The effects of litigation-related stress may trigger emotional distress and other symptoms included in the FBS item pool, so that elevated FBS scores may be an outcome of participation in a lawsuit as opposed to being indicative of malingering (Tsushima and Tsushima 2001).

FBS studies of head trauma patients have most often focused on litigating vs. non-litigating samples while failing to control for the effects of brain injury severity. Yet, brain injury severity is a critical independent variable that affects item endorsements on the FBS and other measures of symptom complaints (Leininger et al. 1991; Martens et al. 2001; Miller and Donders 2001). Head injury severity is inversely related to MMPI-2 scores in litigants and non-litigants, with more severely injured patients commonly scoring within normal limits on the clinical scales (Hoffman et al. 1999). The mild head-trauma victims who participate in these clinical studies typically endorse more FBS items than moderate to severe brain-injured patients, even in the absence of a compensation-seeking context or evidence of malingering (Youngjohn et al. 1997). Perhaps the best explanation for this is that more severe brain injury often impairs individuals' self-insight, and awareness (anosognosia) or appreciation (anosodiaphoria) of their deficits. It is not surprising that, on self-report inventories, these individuals deny or minimize their problems. Second, data suggest that mild traumatic brain injury (TBI), which is far more common than more severe injury, triggers significant psychological reactions in a very small minority of these cases.<sup>1</sup> For these reasons, the relatively common research practice of using moderate to

<sup>1</sup> It is estimated that about 85% of all treated TBIs are mild and that fewer than 1% of mild TBI cases eventuate in a post-concussion syndrome (McCrae 2008), although there are higher estimates, as well.

severe TBI patients as controls for mild TBI litigants constitutes a serious methodological limitation.

There is ample evidence suggesting that TBI and other physical injuries trigger a high rate of somatic complaints and physical preoccupations in a select group of people who are predisposed toward somatoform symptomatology (Greiffenstein and Baker 2001; Kay et al. 1992; Putnam and Millis 1994; Wood 2004). The meaning of high FBS item endorsements in these types of cases is unclear, yet it would seem inappropriate to uniformly categorize them as “faking bad.” For example, there are many reasons that lead to symptom exaggeration and preoccupation that fall short of malingering. In this regard, what is known about the pre-morbid personality characteristics of compensation-seeking mild TBI patients?

In a relevant study, Greiffenstein and Baker (2001) investigated pre-injury and post-injury MMPI-2 profiles in 23 compensation-seeking patients with mild cranial/cervical injuries. These patients attributed major personality changes to their injuries. Their MMPI-2 profiles, before injury or litigation, were abnormal, and their modal code-type was consistent with somatoform disorders. Greiffenstein and Baker (2001) found that the post-injury MMPI-2 profiles of the compensation-seeking mild TBI claimants showed a continuation of somatization trends but, unexpectedly, a general *decrease* in MMPI-2 indices of psychopathology and an increase in defensiveness. Collectively, these findings suggest that compensation-seeking behavior in mild TBI is less related to “faking bad” than to a pre-morbid preoccupation with physical health status and a tendency to develop physical symptoms in response to emotional stress and psychological conflict. Therefore, such symptom endorsements cannot be readily dismissed as “noncredible,” much less as “faking bad,” in all such scenarios. The subtle uniform equation of endorsement of somatoform symptoms with malingering is specious, given current knowledge regarding pain perception and its integral association with psychological processes and attentional mechanisms (Price 1999; Price and Bushnell 2004).

Another area that presents difficulties to the validity of the FBS concerns a possible sex bias. In general, women are more likely to endorse items about somatic complaints compared to men. For example, higher endorsement frequencies for women on the FBS items about headaches (no. 40; no. 176) are apparent in the MMPI-2 manual; and the item about weekly hot flashes (no. 44) was endorsed by 13% of normative women compared with 2% of men, even though only 28% of the women were between ages 40–59 (Butcher et al. 2001).

In another example, concerning face validity, the FBS does not fare well in terms of item selection. One item, about excessive alcohol use (no. 264), if endorsed true, earns one point on another validity scale of the MMPI-2, the F scale (a well-validated MMPI-2 measure of possible

malingering based on rare or infrequent diverse symptoms); however, when that same item is endorsed in the opposite direction, the person earns one point on the FBS. It is a clear indication of a problematic FBS item because, either way the person responds, it is included on one malingering scale or another, and moreover, the manner in which the item is scored on the FBS appears to lack face validity with respect to the intended purpose of the scale; the item relates more to “faking good” than faking bad, if anything.

Item 264 is on the F scale because of empirical verification of its infrequency in the original MMPI normative sample (i.e., 4% of women and 9% of men responded true); in the contemporary normative sample, 77% of women and 56% of men responded false (Butcher et al. 2001). The differential endorsements of men and women in the contemporary normative sample in the scored direction (i.e., a false response is scored on the FBS) provide another example of a possible item bias against women. Moreover, the scoring anomaly of having item 264 on two measures of malingering, but scored true on one (i.e., F) and false on the other (i.e., FBS) came about because Lees-Haley et al. (1991) decided to include a series of “fake good” items on their malingering scale (representing 35% of its content). However, we question the logic in adopting this set of items into a scale purported to measure malingering; the dynamic underlying the “faking good” items does not equate with a motivation to malingering, which is why there are separate MMPI-2 validity scales for exaggerated responding or possible malingering (i.e., the F family) and defensiveness (i.e., L, originally called the Lie Scale, a measure of naïve defensiveness; K, also referred to as the Defensiveness Scale; and S, the Superlative Self-Presentation Scale).

Involvement in adversarial situations can increase the tendency for some individuals to minimize personal faults and deny deviant attitudes and behaviors, a response style captured by the MMPI-2 validity measures of defensiveness (Greiffenstein and Baker 2001; Pope et al. 2006). In addition to personal injury evaluations, other types of assessments (e.g., child custody cases, parole evaluations, employment screening) involve demand characteristics for individuals to present themselves favorably. Butcher and Han (1995) developed the S scale to assess such defensive responding, and eight FBS items are also on S, along with one on L and two on K, other well-validated measures of defensive, as opposed to malingered, responding on the MMPI-2.

In summary, the item bias reflected in the FBS suggests that a number of factors other than malingering contribute to endorsements of its items (e.g., genuine physical problems, psychologically based physical symptoms, positive self-presentation, posttraumatic stress symptoms, or stress related to litigation). The lack of empirical verification of the 43 items selected by Lees-Haley, including examination of the items’ performance across broad

categories of people, argues against its widespread dissemination. Certain groups (e.g., women, persons with disabilities or physical illness, individuals exposed to highly traumatic situations, and those motivated to present themselves favorably) may be more likely to endorse FBS items for reasons other than malingering.

### An Examination of the FBS Criterion Groups

Beginning with the first FBS study, the assignment of people to malingering status was often a judgment call made by the author. The criterion groups of malingerers and nonmalingerers used in the original development of the FBS were described vaguely, without sufficient details for adequate replication, and taken exclusively from the private practice of its first author (Lees-Haley et al. 1991). Other than gender and mean age, no other demographic information, including preexisting conditions or disability status, was provided, nor were there descriptions of the participants' injuries (either real or feigned), or their relevant medical and psychological histories. The types of patients referred to Lees-Haley, as well as the referral sources, were not described, so that we do not know how representative his convenience sample was relative to the general population of litigants in personal injury.

Larrabee (1998) recommended using the FBS to identify "somatic malingering" of brain injury claimants in research based on a sample of 12 patients he selected from his practice records, six of whom completed the original MMPI and six of whom completed the MMPI-2. Only two patients were women, ages ranged from 23 to 59, and education level ranged from 9 to 14.5 years. No information was provided on ethnicity, disability status, or other preexisting conditions, or other demographic characteristics. All 12 patients, according to Larrabee (1998), were malingerers (i.e., base rate of 100%), given their results on four measures of symptom validity testing and the information that there were no significant findings from any medical test, even though all 12 were in litigation claiming brain injury. No information was provided about the underlying incidents that led to the claims. There was no comparison or control group in the study, but Larrabee (1998) concluded that the FBS correctly identified "somatic malingering" in 11 of the 12 cases, which, according to him, is a significantly more accurate performance than the results provided by the F scale, which "only" identified three of the cases as malingerers, using a *T* score of 70 or higher. However, a 70 *T* score is not a recognized cut-off for malingering on F; *T* scores between 70 and 89, inclusive, indicate valid MMPI-2 protocols with a possible "cry for help" response set (Butcher et al. 2001).

More recent studies classify subjects as malingering self-reported somatic symptoms by using various measures

labeled "symptom validity" (e.g., Wygant et al. 2007) that actually assess memory performance (e.g., Test of Memory Malingering; Word Memory Test), not exaggerated responding on personality instruments like the MMPI-2. However, a National Academy of Neuropsychology position paper explicitly states, "Invalid performance on a measure of personality does not allow for an a priori conclusion that the neurocognitive test results are also unreliable, and vice versa" (Bush et al. 2005, p. 424). Clearly defined, objective procedures for assuring that purported malingerers studied are actually faking their reports of somatic or psychological symptoms are lacking in the recent and past research. In addition, examination is lacking of the FBS scale's use with population subgroups like people with disabilities, women, different ethnic groups, the elderly, and so forth. Given the subjective or insufficient basis used for judging a client to be malingering in many research studies in this area to date, their results are not likely to be reproducible. When studies do include validated personality measures of exaggerated responding to MMPI-2 items (i.e., the F family of validity scales and indicators), FBS proponents typically explain the discrepancies with FBS by challenging the validity of the F family and proclaiming that the FBS is a more accurate measure of malingering, as we will see below.

Aside from difficulties in item selection and the research undertaken, the underlying assumption of the FBS can be questioned. The FBS could be aptly renamed the Litigated Personal Injury Scale because its research support is largely based on discriminating between litigating personal injury cases from non-litigating "controls." In many research studies, an underlying assumption has been that personal injury litigants are more likely to consciously fake symptoms. Compensation-seeking examinees are presumed to score higher on FBS because they are typically motivated to lie and fabricate symptoms for material gain (e.g., Fox et al. 1995; Lees-Haley et al. 1991; Tsushima and Tsushima 2001). A plausible alternative explanation is that many people who have injuries that result in physical and emotional problems legitimately pursue compensation via litigation. Moreover, there are other possible response biases in psychological testing other than motivation to malingering; one should not automatically rule in this motivation based on partial, incomplete, or flimsy evidence.

Although Greiffenstein et al. (2004, p. 578) acknowledge in their widely cited study supporting the FBS, "The presence of litigation should not automatically diminish the believability of complaints in persons experiencing extraordinary events," litigation status was used as one of their four criterion measures for including patients in a group they labeled "improbable" posttraumatic syndrome. Greiffenstein et al. (2004) examined FBS scores in three groups drawn from four settings: clinical patients with probable posttrau-

matic syndrome (Prob-PTS;  $n=15$  men, 33 women); a major trauma and litigation group (Lit-MPTS;  $n=11$  men, 21 women); and an improbable posttraumatic syndrome group (Imp-PTS;  $n=26$  men, 31 women).

The criteria for inclusion in this study were documented exposure to a traumatic event, presence of psychological symptoms attributed to the event, and known litigation status (Greiffenstein et al. 2004). Litigation status, time of posttraumatic event, and prescribed psychotropic medications (i.e., narcotic analgesics, antidepressants, sedatives and hypnotics, stimulants and xanthines, and mood stabilizers) were the primary objective indices that differed across the groups. Information about the presence/absence of pre-injury mental health treatment was collected, but not provided. The Prob-PTS group was not involved in litigation, and was a mean of 5 months posttraumatic incident, with most (69%) seen within the first 3 months. The other two groups were involved in litigation and were over 2 years post-incident. The two litigating groups were being treated with significantly more psychotropic medications than the non-litigating “Prob-PTS” group.

Greiffenstein et al. (2004, p. 576) may have minimized the trauma experienced by the “improbable” PTS group (i.e., they were “seeking compensation for psychological damages blamed on relatively minor events”). Other criteria used to classify these patients as “improbable” PTS included, in addition to being in litigation: prolonged disability over 1 year (this could be due to a valid disability status, which is consistent with this group being treated with the greatest number of psychotropic medications) and the presence of “at least two atypical history features,” from a subjectively determined list of nine that Greiffenstein et al. (2004, p. 577) suggest are atypical features of PTS or PTSD. These atypical features developed by the authors to assign litigants to the “improbable” PTS category include, as examples, comorbidity with Axis I or Axis II diagnoses; delayed symptom onset >1 month; dual claims of mild traumatic brain injury and PTSD; and symptoms not responsive to psychotropic medications. However, Greiffenstein et al. (2004) presented no information about the patients’ actual diagnoses (or how they were made) that led to the patients being classified as “improbable” PTS. Some of their suggested atypical features relate to PTSD diagnoses and are open to challenge (e.g., co-morbidity is not improbable; mild TBI can co-occur with PTSD). Furthermore, there was no evidence that the members of the Imp-PTS group reached the criteria for a PTSD diagnosis, as opposed to other Axis I or Axis II diagnoses, and the prescribed medications for their patient groups were broad and included medications used to treat several disorders, not just PTSD (e.g., depressive disorders, significant pain complaints, sleep disturbances, generalized anxiety).

The MMPI-2 was administered to the three patient groups, and the only exclusionary criteria used was Cannot Say scores greater than 30; none of the sample met that criteria. However, four invalid profiles (three with VRIN scores greater than 80  $T$  and one with both VRIN and TRIN greater than 80  $T$ ; the VRIN scale measures response inconsistency, and TRIN is a measure of endorsing either true or false to the items regardless of content) were left in the analysis sample, against best practice (the three invalid profiles based on VRIN were included in the Imp-PTS group, and the fourth was included in the Major-Lit group even though both VRIN and TRIN were elevated). Therefore, the impact of random responding on FBS scores in the litigating groups was neither controlled nor accounted for, despite the fact that random responding does not reflect malingering and generally leads to very high scores on measures of symptom exaggeration.

Each of the three patient groups produced valid MMPI-2 profiles based on the F family scores. FBS raw scores, on the other hand, were elevated, ranging from 20 to 30 for women in the “improbable” PTS group showing the highest elevation, followed by the other litigation group, with undeniable major trauma (i.e., Lit-MPTS). The non-litigants (i.e., Prob PTS) scored the lowest. The same pattern across the three groups was shown for men, with their FBS raw scores ranging from 20 to 28, generally consistent with the pattern of men producing lower FBS scores than women.

The Greiffenstein et al. (2004) study can be criticized in several respects. Its weaknesses include inadequately specified criteria for classifying comparison samples, systematic sampling bias in data collection, inadequate reporting of relevant data, and insufficient data interpretation. First, the comparison samples were poorly defined and did not include Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV; American Psychiatric Association 1994) diagnoses. The context for the MMPI-2 examination of the “probable” PTS group is not clearly specified (e.g., were they self-referrals; part of victim’s or employee assistance programs; subjects recruited for research?). The authors state that these subjects experienced “undeniably frightening events” (p. 578), although that statement is presented without objective support. Similarly, the authors’ characterization of their “improbable” PTS sample as having experienced only a “minor scare” (p. 573) seems to be based on assumptions about these individuals’ emotional status at the time of the trauma. Research psychologists who use archival data are generally incapable of forming accurate retrospective analyses of their subjects’ emotional state at the time of a past incident.

The “improbable” PTS cases had head trauma of ambiguously described severity. Use of the term “benign head trauma” to describe the majority of this sample is questionable in the context of undocumented measures of

brain injury severity (Greiffenstein et al. 2004, p. 577). Substantial subjectivity was involved in classifying subjects, allowing for the methodological threat of investigator bias. For example, the authors had leeway to classify a victim of a high-speed motor vehicle accident who sustained broken ribs as an improbable PTS case, while classifying robbery victims as “probable PTS.” In addition, the authors used the presence of co-morbid diagnoses to classify individuals as “improbable” PTS and symptom exaggerators, without any description of how these diagnoses were formulated, or who made them.

Second, this study exemplifies the sampling bias that characterizes the vast majority of studies of FBS. Most FBS studies fail to control for litigation status or selection bias associated with brain injury severity. Greiffenstein et al. (2004) investigated “trauma claims,” but failed to control for *type* of trauma claims (physical vs. emotional trauma). The “improbable” PTS group was predominantly a physical injury sample. The “probable” PTS group was described as having emotional trauma in the absence of physical or head injury. This distinction is highly significant because, unlike  $F$ ,  $F_b$ , and  $F_p$ , FBS scores are largely based on responses to physical symptom items. It is hardly surprising that brain concussion patients would score higher than physically uninjured robbery victims on a measure that primarily reflects physical symptom reporting (FBS). These results are interpreted as support of the FBS over the other  $F$  scales, contradicting those of better-controlled studies of the  $F$  scales (e.g., Berry et al. 1995; Dearth et al. 2005).

Third, Greiffenstein et al. (2004) failed to report important details of their investigation that are critical to evaluating the validity of their conclusions. For example, they did not publish their information regarding prior psychiatric treatment across groups. Furthermore, they did not report their samples’  $F_b$  scores, which are highly sensitive to emotional trauma and the acknowledgement of problems related to depression and anxiety. Finally, no information was provided on the groups’ performance on the MMPI-2 clinical, content, or other measures of psychopathology that would relate to psychological symptoms present in the three groups.

Fourth, their data interpretation is incomplete because it largely ignores the fact that their sample of litigating bona fide major trauma victims (Lit-MTPS) had a plausible symptom presentation (using the authors’ highly restrictive criteria), yet the FBS misclassified many of them as “faking bad.” Litigants who experienced major trauma and were judged to have “probable” PTS produced a mean raw score of 24.3 on FBS with a SD of 5.7. “Among females, roughly, one half scored above the recommended cutting score of 26 or greater, and two thirds of the males did” (Greiffenstein et al. 2004, p. 583). The authors largely dismiss their own data regarding this issue; that is, it seems

clear that individuals who most likely were presenting with valid symptoms in their study were attributed negative motivations to fake bad by the inappropriate use of the FBS and its suggested cutting scores. Nevertheless, these data by Greiffenstein et al. are used to support the value of the FBS in determining faking bad profiles and suggest that FBS scores now included in the MMPI-2 scoring materials may influence clinicians to classify bona fide victims of posttraumatic stress who are involved in seeking compensation as “faking bad”.

### Reliability and Validity of the FBS

Limited psychometric information on the FBS was available in the first decade of its use, and its psychometric properties require further study. Butcher et al. (2003) provided some psychometric data on the performance of the FBS in several large samples of psychiatric inpatients ( $n=6,731$ ), VA inpatients ( $n=901$  men), chronic pain patients ( $n=4,408$ ), general medical patients ( $n=5,080$ ), inmates in a correctional facility ( $n=2,897$ ), and a smaller personal injury litigant’s sample drawn from nine forensic practices ( $n=157$ ). In all samples, the FBS showed low to modest alpha coefficients, a reliability measure of internal consistency (range 0.43 to 0.74 for women; 0.49 to 0.65 for men), except in the personal injury sample, where somatic symptoms were prominent (alphas were 0.85 for women; 0.86 for men). Other reliability measures (e.g., test–retest) are lacking. In addition, not surprisingly, given its item content, Butcher et al. (2003) found the FBS to be most highly correlated with raw scores on the clinical scales, Hypochondriasis (Hs or scale 1), Depression (D or scale 2), and Hysteria (Hy or scale 3); and the content scales, Health Concerns (HEA) and Depression (DEP). This suggests that FBS appears to be a measure of general maladjustment and somatic complaints, as opposed to malingering.

Furthermore, in the study, the observed correlations between the FBS and the traditional MMPI-2 measures of symptom exaggeration that are based on rare or infrequently endorsed items (i.e., the  $F$ ,  $F_b$  and  $F_p$  scales) were relatively low (i.e., ranging between 0.02 and 0.39 in the five largest samples) compared to their relationships with the clinical and content scales (Butcher et al. 2003). The FBS correlations with  $F$  and  $F_b$  (a measure of infrequency in the back half of the instrument’s items) were also lower (0.53 and 0.55, respectively) in the personal injury sample relative to the relationships of FBS with Hypochondriasis (0.81), Depression (0.84), and Hysteria (0.85). Similarly, Guez et al. (2005) found nonsignificant relationships of FBS scores to the  $F$  and  $F-K$  indices. Proponents explain the lack of relationships with established MMPI-2 measures of symptom exaggeration by suggesting the FBS measures

a different type of malingering than does the F family, one that is especially likely in forensic evaluations for personal injuries (e.g., Greiffenstein et al. 2007; Larrabee 1998, 2007; Lees-Haley et al. 1991).

In addition, Greiffenstein et al. (2007) are highly critical of the F scales, suggesting, “There is a surprising dearth of research on the F family’s validity in natural samples” and the assumption underlying the F family is that “all malingering is alike” (p. 212). However, the MMPI F scale has been researched in both experimentally controlled simulation research (e.g., Bagby et al. 1995; Elhai et al. 2000; Graham et al. 1991; Hunt 1948) and in “real world” studies as well, including correctional settings (e.g., Grossman and Wasylw 1988; Moskowitz et al. 1999; Wasylw et al. 1988); compensation cases (e.g., Charles 2000); inpatients (e.g., Gynther and Petzel 1967; Post and Gasparikova-Krasnec 1979); and with suicidal patients (e.g., Rissmiller et al. 1998), to mention only a few. Furthermore, like F, the FBS was developed using the MMPI item pool initially to identify those “simulating or exaggerating emotional distress” from those “who are not malingering” (Lees-Haley et al. 1991, p. 203). It was only later that FBS was recommended for identifying “somatic malingering” of brain injury claimants (e.g., Larrabee 1998) based on studies with questionable methods of identifying malingering status (see above section on criterion groups). Suggestions that the lack of demonstrated relationships between the FBS and F scales support the discriminant validity of the FBS (e.g., Greiffenstein et al. 2007) are unfounded, given that the FBS was initially developed to measure, albeit within the narrow setting of litigation for personal injuries, a construct closely related to those measured by the F scale.

Unlike the FBS, the F scale was developed empirically and purposely included extremely diverse content “ranging from bizarre sensations, strange thoughts, and peculiar experiences to feelings of alienation and isolation from family members, from others, or from social institutions, or to atypical attitudes towards laws, religion, or authority and to a number of unlikely or contradictory beliefs, expectations, and self-descriptions” (Dahlstrom et al. 1973, p. 114; The three religion items and one item on diarrhea were dropped in the MMPI-2 because that content was objectionable in many contemporary settings.) Next, only items endorsed infrequently in the original Minnesota normative sample (i.e., no more than 10% of the sample endorsed the item in the scored direction) were included on the F scale, based on the premise that only individuals trying to exaggerate or malingering psychopathology will endorse items from broad and inconsistent problem areas that are in excess of what most patients would endorse and do not represent actual syndromes or disorders (Butcher and Williams 2000). There are several widely accepted

possible explanations related to a high elevation on the F scale including: random responding, high true endorsement, exaggerating complaints, confusion, asserting a plea for help, or severely disturbed symptom presentation, and a number of MMPI-2 validity scales (e.g., VRIN, TRIN, F<sub>b</sub>, F<sub>p</sub>) and extra-test variables (e.g., reading problems, cultural factors) can be used to refine its interpretation (Butcher and Williams 2000; Graham 2006). Arbisi and Ben-Porath (1997, p. 102) pointed out that, “Items included in the F scale were also selected to cover a variety of content so that it would be unlikely that any particular set of problems would produce an elevated score on F scale.”

To review, F has been subject to careful development, there are plausible interpretive hypotheses for elevated F scores, and there are multiple research studies on its validity, a situation which contrasts sharply with the narrower set of items selected for the FBS, items that are based solely on one rater’s opinion without any empirical examination of item frequency data or performance across settings. As we have discussed, FBS items are not diverse and primarily overlap with established measures of defensive (i.e., not exaggerated) responding, as well as with symptoms of syndromes measured by the clinical scales Hypochondriasis and Hysteria, the content scale Health Concerns, and many actual physical disorders. Its lack of association with the F family, combined with its association with scales measuring somatoform disorders and depression, are not supportive of claims of it being a more valid measure than the F scale for somatic malingering.

FBS is significantly related to the new Response Bias Scale (RBS) intended to detect over reporting in forensic neuropsychological and disability assessments with the MMPI-2 (Gervais et al. 2007; Nelson et al. 2007). Its developers claim the RBS consistently outperforms the FBS, F, and F<sub>p</sub> scales, citing as evidence the enhanced ability of the RBS in detecting poor performance on one of three cognitive tasks: the Word Memory Test, the Computerized Assessment of Response Bias, and/or the Test of Memory Malingering (Gervais et al. 2007). However, the vast majority of subjects used to develop the RBS did not have head injuries, and as discussed above, poor performance on a memory test should not be used as an indicator of an invalid response set on a personality measure like the MMPI-2 (Bush et al. 2005). Other issues that need to be addressed regarding the RBS are similar to those described in this paper. Especially of concern are false-positive rates when persons with legitimate head injuries, and resulting somatic symptoms, are mislabeled as malingering.

False positives are an identified threat to the validity of the FBS (e.g., Butcher et al. 2003; Iverson et al. 2002). In this regard, a study of whiplash victims is illustrative of the problems that the FBS has in meeting accepted empirical standards of construct validity. Guez et al. (2005) examined

neuropsychological functioning and MMPI-2 responses in 42 patients with chronic neck pain: 21 patients with whiplash trauma (70% had resolved compensation claims) and 21 patients without trauma, contrasting each with a healthy reference group ( $n=500$ ). Malingering was evaluated using effort testing, level of neuropsychological performance, and scores on the MMPI-2 validity scales. Both patient groups had neuropsychological performance levels similar to the reference group (i.e., there was no evidence of impairments in either patient group). With the exception of scores on FBS, there were no MMPI-2 scores suggestive of symptom exaggeration. However, the whiplash trauma group had higher FBS scores than the pain group without trauma, and both patient groups had higher FBS scores than the reference group, in the absence of any evidence of symptom exaggeration in the neuropsychological battery. Any suggestions that FBS possesses better test sensitivity than the MMPI-2 F and F-K validity measures are offset by the lack of supporting evidence for symptom exaggeration.

The results of Guez et al. (2005) suggest that a significant proportion of medical patients who score high on FBS are not malingering. Surprisingly, Nelson et al. (2006; p. 43), in their meta-analysis of research studies on the FBS, mischaracterized the non-malingering whiplash sample used by Guez et al. (2005) as “having reasons to exaggerate symptoms.” The non-malingering whiplash sample was subsequently excluded from the meta-analysis as a “statistical outlier” (Nelson et al. 2006). Had the Guez et al. (2005) data set on the FBS been incorporated into this meta-analysis (i.e., subjects categorized as non-malingers), it would have *lowered* the overall effect size found in the study.

In the context of neuropsychological settings, aside from its other problems that we have documented, a significant amount of evidence suggests that the FBS is neither an unambiguous nor a specific measure of “faking bad.” In common parlance, the term “faking” implies a conscious intent to deceive. However, FBS provides no information regarding an examinee’s conscious intentionality, which is at the very heart of the definition of malingering. Malingering is a conscious and deliberate attempt to obtain some form of advantage or material gain. Somatoform disorders encompass symptoms and behaviors that are neither conscious nor deliberate, and yet are associated with higher MMPI-2 scores on the FBS. There is no evidence that FBS differentiates between malingering (deliberately faking symptoms) and the unconscious expression of psychological conflict through an emergence of physical symptoms and development of health preoccupations (Greve et al. 2006). This limitation is particularly problematic in the assessment of post-concussion syndrome because of its linkage with psychological maladjustment

and associated somatic complaints (Greiffenstein and Baker 2001; McCrae 2008; Youngjohn et al. 1997).

There are a number of other research studies with results that do not support the validity of the FBS scale for identifying malingerers. Bury and Bagby (2002) compared people with work-related injuries, who met DSM IV (1994) criteria for PTSD, based on a structured clinical interview, with a group of normal individuals who were asked to fake the MMPI-2 under varying conditions. The faking groups scored higher on all the invalidity measures on the MMPI-2 except FBS. The authors found that the FBS was ineffective at detecting fake-bad responding in personal injury cases.

Bury and Bagby’s (2002) study was criticized because it did not include other measures of symptom validity besides those on the MMPI-2 in their PTSD sample seeking compensation for workplace injuries (e.g., Greiffenstein et al. 2007; Larrabee 2007). However, the symptom validity or criterion measures used in FBS studies described by Larrabee (2007) are not measures related to psychological or somatic symptom exaggeration (i.e., the content covered in the MMPI-2 and measured by the MMPI-2 F family); rather, they are performance measures related to memory functioning and, as such, are not adequate criterion measures for exaggerated MMPI-2 responding, as we have mentioned (Bush et al. 2005). Furthermore, in addition to a structured clinical interview for determining the subjects’ PTSD diagnoses, Bury and Bagby (2002) documented the subjects’ workplace-related traumas as follows: physical assault (23%); transportation accident (16%); pinned by machinery (12%); electrocution, fire, or explosion (12%); witnessing another’s death (8%); and other trauma (13%), which are all consistent with DSM-IV (1994) diagnostic criteria for PTSD; such consistency between self-report and documented history is a common method for assessing symptom validity (Bush et al. 2005).

Burandt (2006) conducted a study of the effectiveness of the FBS and concluded: “therefore, there was no significant difference in FBS scores among the electrically injured patients who failed or passed symptom validity testing. Based on these findings, the FBS is not yet a reliable tool for detection of somatic malingering in electrical injury.” Arbisi et al. (2006) found that the FBS did not add incrementally to the prediction of exaggerated PTSD. Other studies demonstrate that FBS scores are elevated in persons with objectively verified sources of physical discomfort (Iverson et al. 2002; Meyers et al. 2002). Pre-morbid psychiatric history is associated with higher scores on FBS (Martens et al. 2001). Scores on FBS can be raised significantly by the presence of acquired neurological deficits, including anosmia and motor impairment in brain-injured patients (Greiffenstein et al. 2002). The data cited suggest that self-reports of bona fide physical

maladies combined with psychological distress can produce high scores on FBS.

Two meta-analyses that compared the FBS with other MMPI-2 indices of exaggerated responding reached different conclusions about its validity (Nelson et al. 2006; Rogers et al. 2003). Rogers et al. (2003) used 65 MMPI-2 feigning studies plus 11 MMPI-2 diagnostic studies reporting that the MMPI-2 F scale was supported as an effective scale for detecting malingering. They found the FBS to be generally ineffective in assessing malingering because of its highly specific and narrow focus. In contrast, Nelson et al. (2006) found that clients in litigation report a high number of physical symptoms, which they labeled “overreporting.” However, these authors did not demonstrate how they differentiated “overreporting” from “reporting” of legitimate symptoms. Furthermore, Nelson et al. (2006) minimized the results of the earlier meta-analysis because Rogers et al. (2003) relied upon peer-reviewed publications. Use of peer-reviewed publications in a meta-analysis is a widely accepted standard. In contrast to this standard approach of relying upon published (and readily available) results, Nelson et al. (2006) included several samples from unpublished sources and, as noted earlier, the data in Guéz et al. (2005) were excluded, along with data from several other published studies. The above discussion about the procedures used in the competing meta-analyses of the FBS suggests that Rogers et al. (2003) used more defensible methodology upon which to reach their conclusions.

Other summary evaluations of the FBS questioned its utility. Bagby et al. (2006, p. 69) concluded that the scale is too problematic to be used in assessment decisions: “The assessor should be aware that it is highly probable that the FBS will classify an inordinate number of disability claimants with bona fide health and psychiatric problems as malingerers.” Graham (2006, p. 39) recently reviewed the research on the FBS and recommended: “the FBS not be used to identify malingering of psychopathology on the MMPI-2.” Also, Arbisi (2006, p. 418) concluded that “Consequently, results from these studies as well as from the Rogers et al. meta-analysis indicate that the FBS should not be used as a means of identifying individuals who are faking-bad on the MMPI-2 regardless of setting.”

Group studies that average test scores over many individuals indicate that seeking compensation is associated with a slightly lower level of neuropsychological test performance (Binder and Rohling 1996). However, these results are not consistently replicated, and even if they proved replicable, there are multiple reasons that do not involve malingering that could be used to explain lower neuropsychological scores in populations with psychological injury such as the effects of concomitant pain, headaches, and sleep difficulties (Nicholson and Martelli 2007). MMPI-2 indices of depression and anxiety are associated

with performance decrements in TBI cases (Gass 1996; Ross et al. 2002). In this regard, in their TBI sample, Ross et al. (2002) also found that MMPI-2 indices were more powerful predictors of performance on attention and memory tests than was litigation status. There are additional explanations for findings related to symptom exaggerations that do not include malingering (e.g., the effects of a cry for help). Furthermore, this apparent association between compensation-seeking status and increased symptoms is often interpreted as indicative of a unidirectional causal relationship (viz., compensation seeking causes a person to fabricate symptoms), despite the well-known admonition by statisticians to avoid confusing correlation with causality. As the influential British empiricist David Hume elucidated quite clearly in the 18th century, this specific type of confusion is pervasive in human thought. For example, guidelines for using correlation to infer causality were cited by Larrabee et al. (2007). Their two fundamental criteria—strength of correlation and regularity of co-occurrence across populations—are examples of the confusion. Based on these criteria, one could illogically conclude that diverse populations of people across the world, waking up, and eating breakfast cause the sun to rise and the outdoor temperature to steadily increase. Concerning the specific examples related to compensation seeking, one could argue that any sign of increased symptoms and lowered neuropsychological performance are the reasons for the compensation seeking, rather than the compensation seeking being the reason for the altered presentation. Clearly, careful research is needed on the question rather than facile conclusions.

### Evaluating FBS Responding: Raw Score Cut-Offs and Norms

The development of norms is essential for the interpretation of scores on high stake assessments like the FBS, for they let the test user know, for example, for every possible test score, what percentage of the entire population earned scores in that range and below (Geisinger 2005). All of the MMPI-2 validity scales are interpreted using standardized *T* scores (Butcher et al. 2001). Interpretation of the FBS has not followed this traditional approach relying, instead, on various raw score cut-offs that range from the originally proposed 20 to a high of 30 plus, despite its developers’ observations early on that “a percentile score or comparable standard score is often more informative than a single dichotomous cut-off (Lees-Haley et al. 1991).

Rather than presenting information on the routinely used *T* scores for MMPI-2 interpretation, Greiffenstein et al. (2007) presented FBS raw scores as standardized *z* scores using the MMPI-2 normative sample. They concluded that

FBS raw scores of 17 and 19 for men and 20 and 22 for women represent the 93rd and 98th percentiles, respectively; and, based on *z* scores, concluded that “elevated scores on the FBS are rare among community residing persons not claiming mental health disorders” (p. 215). No explanation was provided for using *z* scores instead of *T* scores, or how the findings inform an interpretive strategy other than individuals from the MMPI-2 normative sample produce lower *z* scores on the FBS.

Although the new FBS scoring materials include profile sheets that plot FBS raw scores as standardized *T* scores, the MMPI-2 publisher and distributor recommend reliance on a raw score range of greater than 22 to greater than 28 (Ben-Porath and Tellegen 2007b; Pearson Assessments 2007), with no accommodations made for gender, unlike the case when *T* scores are used, as with other MMPI-2 measures. Greiffenstein et al. (2007, p. 220) provide two generalizations about the use of raw score cut-offs for women: “first, females score higher on the FBS by roughly 2 items. Second, a sizeable minority of females with psychiatric histories score above 20, assuming normal kurtosis.” Although the publisher refers psychologists to the Greiffenstein et al. (2007) article for more information about the FBS, its general recommendations for using the FBS does not take into account Greiffenstein et al.’s reported gender differences, nor does the publisher’s report that gives FBS scores for patients provide the psychologist with basic information about the raw score equivalents of *T* scores at the standard levels of 50, 65, 70, 80, and so forth, to assist in evaluating possible bias against women.

Table 1 presents the *T* score values for the various FBS raw scores for men and women developed by Greene

(2000), illustrating the differences in raw and *T* scores for men and women at various cut-off points. The *T* score equivalent of the recommended FBS raw score of 29 for identifying malingering (Ben-Porath and Tellegen 2007a, b; Pearson Assessments 2007) for women is 87, but the equivalent for men is 95, close to a full SD higher! Furthermore, using guidelines developed for interpreting possible symptom exaggeration based on the F scale in outpatient settings, *T* scores between 70–89 are considered valid, with possible exaggeration of symptoms, perhaps as a “cry for help” (Butcher et al. 2001). Profile invalidity for the F scale does vary depending upon setting: psychiatric inpatient (*T* scores >100); outpatient (*T* scores >90); and nonclinical (*T* scores >80). Nonclinical settings are ones from the general population, like personnel screening. Personal injury evaluations take place in outpatient settings or, in some cases, during or after hospitalizations. Table 1 illustrates that in outpatient settings, the FBS raw score equivalent to a *T* score greater than 90 for women is 31, but only 27 for men, further illustrating our point about the potential for bias against women when following the publisher’s recommendation of using the same raw score cut-offs for men and women.

In contrast to the currently recommended FBS cutting score of 29 for identifying malingering, the initial recommendations to use an FBS raw score of 20 as the cut-off for identifying malingering (e.g., Lees-Haley et al. 1991) meant that essentially normal range scores would lead to misclassification of malingering across all settings, including the nonclinical, and for both genders. An FBS raw score of 23, currently suggested by the publisher as raising concerns about the validity of a person’s self-reported MMPI-2 symptoms, is also a normal range score for women across all settings, using the F scale recommendations that are based on *T* scores as guidelines. For men, an FBS raw score of 23 equals a *T* score of 80, and in clinical settings this score should suggest a cry for help in an otherwise valid profile. The two other members of the F family,  $F_b$ ,  $F_p$ , share similar *T* score cut-offs (Butcher et al. 2001).

Given the relatively recent introduction of *T* scores for interpreting the FBS, and the problems noted above, it seems important to determine how consistent the recommendations are for raw score cut-offs. A variable yardstick is apparent in proposals about how the FBS should be used to identify malingering, with no clear consensus emerging for any of the suggestions from the many proposals. We provide a brief illustration: the initial suggestion was to use the criterion of 20 or higher (e.g., Lees-Haley et al. 1991); and it is set at 20 or higher for those with “mild head injuries,” but it is set at 24 or higher for “moderate/severe head injuries” (Greiffenstein et al. 2002). There is also mention of the criterion of 21 or higher for men and 26 or higher for women (Greiffenstein et al. 2004), and 24 or

**Table 1** *T* score equivalents for raw score cut-off points for the FBS (Adapted from Greene 2000)

Raw score	Men <i>T</i> score equivalent	Women <i>T</i> -score equivalent
20	72	65
21	74	67
22	77	70
23	80	72
24	82	75
25	85	77
26	88	80
27	90	82
28	93	84
29	95	87
30	98	89
31	101	92
32	103	94
33	106	96
34	109	99
35	111	101

higher for men and 26 or higher for women to identify “pseudo-PTSD” (Lees-Haley 1992).

In addition to the lack of consensus regarding optimal cut-off raw scores, studies are quite variable in terms of the base rates of malingering in their samples; as examples, it was 100% in Larrabee (1998) and ranged from 25 to 50% in the table for positive predictive power of the FBS presented in Greiffenstein et al. (2004), with their suggestion that 50% is common in worker’s compensation settings. These high rates are not unusual in FBS studies, yet are well above the 1 to 20% base rate for malingering reported by Sharland and Gfeller (2007) in their survey of practitioners. How well these high FBS base rates for malingering apply across most settings that use the MMPI-2 is questionable, at best, including those involving psychological injury.

Ben-Porath and Tellegen (2007a, b) refer psychologists to Greiffenstein et al. (2007) for further guidance in using their generic recommendations for FBS interpretations of cut-off scores ranging from over 22 to above 28. However, careful inspection of Greiffenstein et al. (2007) reveals that they provide the following variable and imprecise guidelines about identifying malingering using the FBS:

- 23 or higher justifies concerns about symptom validity no matter what the type of patient is,
- 29 or higher justifies concerns about malingering for women with pre-injury psychiatric history,
- the criterion score can be as low as 23–24, for cases with historical or radiological evidence negative for cerebral dysfunction,
- it is 26 or higher in cases with severe brain injury with residual neurological signs,
- it is 30 or higher in cases of serious, active medical disease.

Berry and Schipper (2007) observed that the presence of genuine neurological disorder might encompass symptoms that are reported on the MMPI-2, thereby elevating scores on the MMPI-2 validity and clinical scales. Although this does not appear to be a problem on the L, K, and F scales (Gass 1991; Gass and Wald 1997), brain-injury symptoms are associated with higher scores on FBS in bona fide TBI samples (Greiffenstein et al. 2002). Berry and Schipper (2007) concluded that it might be necessary to derive and validate new cut-off scores for the FBS in such cases. As indicated above, Greiffenstein et al. (2007, p. 228) also recommended that severe brain injuries involving residual neurological signs should lead clinicians to adjust the FBS “cut-score to 26 and up.” At this time, there is no clear agreement or consensus of opinion as to what constitutes optimal cutting scores on FBS across highly variable circumstances and a wide variety of forensic and other neuropsychological examination contexts (e.g., criminal or

civil; neuropsychological or psychiatric evaluation; Berry and Schipper 2007). Although the authors are optimistic about the use of the FBS in civil forensic neuropsychological settings, their cumulative data derived from five studies indicate an average specificity of only 0.76 in neurological samples, with specificities across studies as low as 0.53 (Berry and Schipper 2007). That is, their obtained proportions of erroneously labeled individuals, or false positives, is unacceptably high, given the potential for harm to patients if incorrectly labeled as “malingering.” Their data raise serious concerns about the possibility of drawing erroneous conclusions about brain-injured individuals using FBS.

### Illustrations of Potential for Bias with Widespread Dissemination of FBS

To summarize, with the publisher’s decision to include the FBS in the MMPI-2 scoring materials, clinicians must carefully evaluate how this decision will impact the MMPI-2 profiles of the patients they evaluate, and they must avoid applying the pejorative label of malingering or “faking bad” to those patients who are accurately reporting their symptoms or exaggerating for reasons other than malingering. To illustrate this type of difficulty raised by the blanket use of the FBS in MMPI-2 scoring materials, we describe two groups of patients that could be negatively impacted by use of the FBS, given how their responses to the FBS could be elevated, yet not be indicative of malingering: veterans treated in VA inpatient units and women in residential treatment for eating disorders.

Butcher et al. (2003) provided information on the high percentages of mental health patients with scores higher than the various FBS cut-offs, including 901 men hospitalized in a VA psychiatric unit who produced valid MMPI-2 profiles based on standard validity criteria (e.g., Butcher et al. 2001). Depending on the cut-off used, the FBS yielded malingering rates in this setting of 47% (cut-off of 20 or more), 33% (cut-off of 22), or 24% (cut-off of 24). According, then, to the FBS, a substantial number of veterans in a tertiary care residential program are “over-reporting” or malingering. How likely is it that somewhere between 24–47% of VA inpatients were misdiagnosed and inappropriately hospitalized? Up to this point, we have shown clearly that the FBS does not measure what it purports to measure, its cut-off scores are variable and elusive, and its psychometric properties not validly demonstrated. Should psychologists be more skeptical about their patients’ self-reports of psychological and somatic symptoms based on elevated FBS scores but otherwise valid MMPI-2 profiles and performance on neuropsychological evaluations, or should they be more skeptical of the

FBS scores that suggest malingering in such cases? That the VA is seeing an influx of returning Iraqi war veterans with head injuries and PTSD raises concerns of whether these patients will be mislabeled as malingerers, possibly resulting in denial of treatment and/or compensation, given the adoption of the FBS on the MMPI-2 and the blanket assertions that it is a useful measure for identifying feigned disability, such as a more subtle form of non-credible responding in those with mild brain injury (Pearson Assessments 2007).

Interestingly, among their criticisms of Butcher et al. (2003), Lees-Haley and Fox (2004, p. 333) indicated “five of their six samples were not in the setting for which the FBS was designed, and only one—their smallest sample—was personal injury litigants. Applying the FBS in other contexts is analogous to applying the Marital Distress scale to single persons.” As we will see in the next section on the publisher’s statement, now that the FBS is added to the MMPI-2, it is being applied much more broadly, including in the settings described by Butcher et al. (2003), as well as the next one we describe below (see Arbisi and Butcher 2004 for more responses to Lees-Haley and Fox 2004, and Greve and Bianchini 2004 regarding their criticisms of Butcher et al. 2003).

For another population where use of the FBS could result in mislabeling genuine psychological distress as malingering, we present original data to illustrate our point. Women in residential programs for eating disorders are more likely to be characterized as denying problems, rather than exaggerating or malingering (Viglione et al. 2006; Vitousek and Stumpf 2005); therefore, a reasonable hypothesis is that, as a group, they should score below the FBS cut-off scores for malingering. We identified 2,054 women (mean age=26.15 years; SD=8.28; 94% white) with valid MMPI-2 profiles from an initial sample of 2,273 (90% response rate) consecutive admissions to the Remuda Intensive Inpatient Eating Disorders Program between 1/1/2000 and 07/31/2007 and examined their FBS scores. Patients typically are referred to this facility after less intensive programs have been unsuccessful at promoting recovery. At admission, a licensed medical provider, licensed psychiatric provider, licensed psychologist, licensed or registered nurse, registered dietitian, and licensed master’s level therapist assess the patient. All patients completed an extensive battery of psychological tests, including the MMPI-2.

We used the following MMPI-2 criteria to determine valid responding: fewer than 30 omitted items; VRIN *T* scores less than 80; TRIN *T* scores less than 80; F *T* scores less than 100; L *T* scores less than 80; and K *T* scores less than 75 (Butcher et al. 2001). All 2,054 patients met DSM-IV (1994) eating disorder criteria: 37% were diagnosed with anorexia nervosa; 35% with bulimia nervosa; and 28%

with eating disorder not otherwise specified (EDNOS). The vast majority, 97%, were diagnosed with one or more co-occurring DSM-IV (1994) Axis I disorders, with 91% having one or more mood disorders, 79% having one or more anxiety disorders, and 28% having one or more substance use disorders.

Licensed medical providers also diagnosed all patients in our sample having valid MMPI-2 profiles with one or more Axis III medical disorders, with a mean of  $5.8 \pm 2.5$  Axis III disorders. Axis III disorders were typically diagnosed via objective measures, including results of comprehensive laboratory orders standard for all admissions, X-rays, and observable physical signs. Common physical signs included slowed heart rate, electrolyte abnormalities, muscle mass depletion, swollen parotid glands, and multiple objective indicators of malnutrition.

Patients in our sample with anorexia were admitted with a mean body mass index (BMI) of  $15.3 \pm 2.2$ . According to medical standards (Heyward and Stolarczyk 1996), BMIs less than 18.5 indicate underweight, and BMIs less than 15 indicate starvation. The mean BMI for the patients with anorexia, 15.3, neared the starvation level, with 41% meeting the strict starvation criterion of BMIs less than 15. Patients’ mean length of stay in intensive inpatient treatment was  $50.6 \pm 14.0$  days, during which time, they were followed closely by a team of six licensed or registered healthcare professionals. Given the extensive assessment process, objective data sources, documented eating disorder-produced medical compromise of the patients, and 50.6 days of intensive medical and psychiatric monitoring, the likelihood of malingering in this patient population is virtually nil.

Table 2 presents FBS data for this inpatient eating disorder population. The original cut-off score of 20 or higher, used by its proponents during its first decade of use,

**Table 2** Number and percentage of patients with various FBS cut-off scores, by eating disorder diagnosis

FBS raw score	Anorexia <i>n</i> =755 (37%)	Bulimia <i>n</i> =730 (35%)	EDNOS <i>n</i> =569 (28%)	All <i>n</i> =2,054 (100%)
≥20	434 (58%)	451 (62%)	390 (69%)	1,275 (62%)
≥22	352 (47%)	354 (49%)	333 (59%)	1,039 (51%)
≥23 <sup>a</sup>	303 (40%)	306 (42%)	292 (51%)	901 (44%)
≥24	261 (35%)	255 (35%)	259 (46%)	775 (38%)
≥26	179 (24%)	173 (24%)	184 (32%)	536 (26%)
≥28	116 (15%)	92 (13%)	118 (21%)	326 (16%)
≥29 <sup>b</sup>	80 (11%)	66 (9%)	89 (16%)	235 (11%)
≥30	57 (8%)	48 (7%)	64 (11%)	169 (8%)

<sup>a</sup> Cut-off score suggested by test publisher as raising “concerns about the validity of self-reported symptoms.”

<sup>b</sup> Cut-off score suggested by test publisher as raising “very significant concerns about the validity of self-reported symptoms.”

would have misclassified 62% of these severely ill women as malingerers. More recently proposed cut-scores in the range of 23 or higher to 26 or higher would result in misclassifications ranging from 44% to 26%, unacceptably high false positive rates. Using the publisher's recommended cut-off of 29 or higher, 11% of these patients would be classified as malingering, also an unacceptably high false-positive rate, given the independently confirmed health and mental health disorders present in this sample. For example, women with FBS scores of 29 or higher had significantly more diagnoses of Axis III physical disorders ( $6.1 \pm 2.6$ ) than the lower scoring group ( $5.7 \pm 2.4$ ;  $t=2.60$ ,  $p=0.009$ ,  $df=2.052$ ), and virtually all received a co-occurring Axis I diagnosis (99.6% compared with 96.6% of those with lower scores; chi square=6.10,  $p=0.01$ ,  $df=1$ ), in addition to their eating disorder diagnoses (34% anorexia nervosa; 28% bulimia nervosa; and 38% EDNOS). There were no differences in age, length of stay, BMI at admission for those with anorexia, or substance abuse disorders between those with FBS scores above the publisher's cut-off of 29 compared to those with lower scores. Nor were there any indicators from sources other than their "elevated" FBS score pointing to malingering in this group of 235 severely ill women.

Greiffenstein et al. (2007, p. 223) claimed, "Scores of 30+ never or rarely produce false-positive errors. Put differently, the positive predictive power of the FBS >29 is virtually 100%," which led them to their "best practice guideline" (p. 228) of scores of 30+ for the FBS cutting score to identify malingering as having "the greatest confidence irrespective of gender, medical, or psychiatric context" (p. 229). However, Table 2 indicates that 8% of this eating disorder sample that we studied would be classified as malingerers even using this cut-off score of "greatest confidence." Given that the stated purpose of the scale is to identify people falsely claiming injuries or illnesses, thereby preventing them from unjustly receiving financial compensation or recovery of treatment costs, the consequences of this misclassification rate could result in the denial of essential, often life-saving, healthcare interventions in populations of women in psychiatric care for the types of cases that we studied, and presumably, for many others. In fact, using the range of cut-off scores suggested by the publisher, including their most stringent cut-off of >28, the results from our eating disorder population suggest that the specificity of the FBS—its ability to rule out malingering when it is absent—may range from 56% if using a cut-off raw score of  $\geq 23$  to 89% if using a cut-off raw score of  $\geq 29$  among female psychiatric patients. Such specificities would be considered unacceptable by many authorities, as, coupled with low to moderate base rates of malingering, they would compromise the positive predictive power or utility of the measure (e.g., Nunnally and Bernstein 1994).

## Evaluation of Publisher's and Distributor's Statements on the FBS

Instead of a manual or other interpretive guide, the publisher and distributor of the MMPI-2 included on each of their websites a brief statement by Ben-Porath and Tellegen (2007a, b) describing the process for adding the FBS to the MMPI-2 scoring materials, which involved input:

"From a panel of experts on the advisability of adding the FBS to the standard MMPI-2 test materials. Eight experts were asked to review two recent publications that summarize the FBS research base conceptually (Greiffenstein et al., in press),<sup>2</sup> and empirically (Nelson et al. 2006), and to indicate to the University of Minnesota Press whether, in their opinion, the FBS should be added to the MMPI-2 materials."

Ben-Porath and Tellegen (2007a, b) provided the following summary of the eight reviews, "a strong majority of the experts recommended that the FBS be added to... the MMPI-2", as well as the following description of the majority's opinions on how the FBS should be used:

"They agreed that...scores on the FBS should be considered in the context of scores on the other validity scales, the circumstances of the assessment, and any conditions such as significant physical injury or disease that could artificially elevate scores on the FBS. With these caveats in mind, the experts supporting addition of the FBS to the standard set of MMPI-2 validity scales recommended that raw scores above 22<sup>3</sup> should raise concerns about the validity of self-reported symptoms and that raw scores above 28 should raise very significant concerns about the validity of self-reported symptoms, particularly with individuals for whom relevant physical injury or medical problems have been ruled out. More specific recommendations...are provided by Greiffenstein et al. (in press)

Documents provided by the University Press under the Minnesota Data Practices Act<sup>4</sup> elaborate on the decision-making process and recommendations of each reviewer. The documents reveal that in December 2005, Ben-Porath

<sup>2</sup> Cited as Greiffenstein et al. (2007) in reference list.

<sup>3</sup> On the Pearson Assessments' website (i.e., Ben-Porath and Tellegen 2007b) this cut score is 22. However, a verbatim statement appears on the University Press website (i.e., Ben-Porath and Tellegen 2007a) with the noteworthy exception that the recommended lower level cut score is 23, another example of the variable yardstick for cut score interpretive guidelines.

<sup>4</sup> These documents are available as a PDF by emailing the first author at butch001@umn.edu.

and Tellegen received a solicited letter from the developer of the FBS endorsing the MMPI-2RF (an MMPI-2 short form being developed by Ben-Porath and Tellegen at the time) that included his recommendation that the FBS be included in the MMPI-2RF. In January 2006, University Press management asked eight psychologists whether the FBS should be added to the MMPI-2 because “our advisors cite a growing body of research documenting the utility of the FBS as a measure of overreporting of somatic problems and poor effort on cognitive measures.” The reviewers were asked to address two specific questions:

- 1) “Does the literature support the validity of the FBS scale as a measure of overreported somatic complaints and poor effort on cognitive tests?”
- 2) If the scale is added to the standard scoring materials for the MMPI-2, are the interpretive recommendations and caveats provided by Greiffenstein et al. sufficient, or would you recommend changes or additions?”

The following raises questions about the decision-making process to include the FBS in the MMPI-2 scoring materials:

- 1) Two reviewers, the first and second authors of the MMPI-2 manual (Butcher et al. 2001), unequivocally recommended against including the FBS in the MMPI-2. Both cited numerous flaws in the interpretive recommendations by Greiffenstein et al. (2007). However, the documents reveal no direct sharing of this information or any other discussions among the eight psychologists, even though many of the reviews included unpublished analyses and detailed discussions of various psychometric issues.
- 2) One reviewer was asked to revise his first review based on an additional article sent to him by a University Press advisor, although the documents provided by the University do not indicate that the other seven reviewers were also sent this new information for comment.
- 3) Communications from University Press management to the other six reviewers stated “seven of the eight reviewers reported that there is evidence that FBS can function as a validity scale. Of these seven reviewers, there was good agreement among six that the scale should be added to the standard scoring materials for the MMPI-2 and that scores above 28 on FBS will yield very few false positives.” An examination of the content of the six reviews finds more diverse recommendations about appropriate cut scores, consistent with the varied views described above from the literature:

Reviewer 1 “FBS>23 <29. The MMPI-2 profile may be invalid given certain conditions (no

documented pre injury or *post injury* psychiatric illness, no historical or radiological evidence of a cerebral dysfunction, no active medical conditions). FBS>29 the profile is likely invalid due to the malingering or feigning of cognitive deficits.”

Reviewer 2 “if in the interpretive report your company wishes to use language about ‘exaggeration, even possible intentional exaggeration’, there is greater confidence for this sort of language at 26 and above... An even greater level of confidence in terms of making assertions about intentional exaggeration exists when one is talking about scores at 30 or higher”.

Reviewer 3 (*first review*): “Cutting scores above 23 are fairly impractical as false positive rates are at or below 10%, but the sensitivity drops off quickly. A cutting score of 29 (!) has almost no utility at all. Almost no one scores that high.”

Reviewer 4 “In most cases, I would be most comfortable with a cut-off in the high 20s or 30+.”

Reviewer 5 “Malingering becomes are (sic) more probable diagnosis (though does not exclude somatization) as scores rise above 26. Score (sic) of 30 and greater almost exclusively reflect malingering, though, again, the possibility that a somatoform coping style is co-morbid cannot be ruled out and is not unreasonable.”

Reviewer 6 “Thus, an FBS score of 22, in the company of failure on the Word Memory Test and Test of Memory Malingering, is much more likely to be a “true positive” than an FBS score of 25 in a female litigant with a preexisting psychiatric history who is claiming psychological trauma following a physical assault, with no other evidence of symptom exaggeration or performance invalidity. Perhaps some discussion of these issues can be included in the interpretive text, including Greiffenstein’s caveats.”

- 4) Two of the above six reviewers offered substantive criticisms of the Greiffenstein chapter’s interpretive guidelines, whose use was recommended by Ben-Porath and Tellegen (2007a, b):

Reviewer 1 “I do not find the guidelines for the interpretation of the FBS contained in

the Greiffenstein, Fox, and Lees-Haley chapter to be specific enough and sufficiently circumscribed to prevent the misuse of the scale since those guidelines do not provide a framework for the interpretation of the scale across the entire range of settings that the MMPI-2 is applied.”

Reviewer 3 “The Greiffenstein chapter and Nelson paper promulgate many of the problems that occur in the FBS literature, which is a persistent failure to identify the populations of interest and to stick to those populations...”

- 5) Reviewer 2 found the chapter to be “generally on target,” with the following important caveat: “I disagree that 23 is the universal cut-off.” The remaining three reviewers, representing the minority viewpoint of the eight reviewers, recommended use of the Greiffenstein et al. (2007) chapter for interpretations.

With the inclusion of the FBS in the MMPI-2 scoring materials, the risk of harm to patients genuinely suffering psychological distress by unjustly mislabeling them as malingers has been elevated. The reviewers of the decision to add the FBS to the MMPI-2 made disparate recommendations regarding raw score cut-offs. The recommended interpretive guidelines are ambiguous. Given this, it would be hard to argue in a court of law that the evidence supports use of the FBS as a scientifically accepted measure that can be included as part of expert testimony. In the following section, we see, in fact, that three courts refused to allow FBS scores as part of expert testimony based on the MMPI-2.

### **Frye Hearings on FBS in Florida Courts**

For a measure to be admissible in court proceedings, it must be scientifically sound and acceptable in the professional community. This is determined based upon the standards set forth in each state. Florida courts, for example, conduct “Frye” (*Frye v. United States* 1923) hearings when the admissibility of a psychological measure is challenged. There have been three Frye challenges in Florida on use of the FBS to determine malingering; all three resulted in its exclusion. Circuit Court Judge Charles E. Bergmann, in his Frye hearing order for *Williams v CSX Transportation, Inc.* 2007, p. 2, indicated he had to determine if the FBS was “generally accepted in the psychology/neuropsychological community as a reliable assessment of effort and malingering and does it pass the Frye test for admissibility?” He noted that the burden is “to prove the general acceptance of

both the underlying scientific principle and testing the procedure used to apply the principle to the facts of the case at hand” by the side wanting to use the FBS, in this case, the defense.

The plaintiff in *Williams v CSX Transportation, Inc.* 2007 was attacked at her workplace by an unknown assailant who strangled and knifed her repeatedly. She underwent a compulsory forensic neuropsychological evaluation by an expert retained by the defense. The Frye hearing examined whether expert testimony based on the FBS would assist the jury in determining whether the plaintiff had traumatic brain injury or PTSD, and if the plaintiff was malingering, exaggerating, or over reporting her symptoms given her FBS score of 26 in an otherwise valid MMPI-2 profile.

Based on the evidence presented during the Frye hearing, Judge Bergmann (*Williams v CSX Transportation, Inc.* 2007, p. 11) concluded:

“The FBS is very subjective and dependent on the interpretation of the person using or interpreting it. There is no definitive scoring because scoring has to be adjusted up and down based on the circumstances and there is a high degree of probability for false positives. Moreover, the scoring assessment has changed over the years from an original cut score of 20 in 1991, with recommended interpretive scores now ranging from 23 to 30; this coupled with the acknowledged bias against women and those with demonstrated serious injuries makes the FBS unreliable.”

Judge Bergmann (*Williams v CSX Transportation, Inc.* 2007) prohibited the defense expert from testifying in his courtroom that the FBS showed the plaintiff had been malingering. He joined two other Florida judges who also excluded expert testimony about the FBS in cases appearing before them (i.e., *Vandergracht v. Progressive Express et al.* 2005; *Davidson v. Strawberry Petroleum et al.* 2007.)

### **Concluding Comments**

Based on a review and a careful analysis of a large amount of published FBS research, the FBS does not appear to be a sufficiently reliable or valid test for measuring “faking bad,” nor should it be used to impute the motivation to malingering in those reaching its variable and imprecise cutting scores. We agree with the conclusions of the three judges in Florida that the FBS does not meet the Frye standards of being scientifically sound and generally accepted in the field, and that expert testimony based on the scale should be excluded from consideration in court. Throughout this article, we have raised questions about its

potential for bias against women, those with disabilities and physical illness, psychiatric inpatients, individuals exposed to highly traumatic situations, and those motivated to present themselves in a favorable light. The samples used to develop the FBS are not broadly representative of the populations evaluated by the MMPI-2, nor are its criteria used to define malingering objective and replicable. There is insufficient evidence of its psychometric reliability or validity, and there is no consensus about appropriate cut-off scores or use of norms. The publisher's and distributor's statement on its use is not an adequate substitute for a test manual, nor does it convey sufficient cautions about potential misuse of the FBS described in the literature and in many of their reviewer's comments.

All this should give psychologists pause in using the FBS with patients until they have evaluated its underlying research basis. First and foremost, clinicians must ensure that their assessment techniques are based on sound research showing demonstrated reliability and validity, with a clear understanding of how to use the measures so that any potential for harming individuals is kept at the lowest level possible. The potential for misuse of the FBS has increased substantially since the beginning of 2007, when it began to be disseminated widely as part of the standard MMPI-2 scoring. Widespread use in clinical settings could result in many genuinely vulnerable persons being mislabeled as malingerers, with a resulting denial of needed psychological and medical care and/or legitimate compensation for disability. To avoid such outcomes, psychologists must closely compare the research underlying its development with the recommendations for its use.

Several research issues are unresolved at present regarding the clinical use of the FBS. Major concerns include the need for better definitions of the construct the FBS is purported to measure through the use of objective and replicable criterion measures for determining malingering status. It is inappropriate to ascertain that being in litigation, rated by a psychologist as exaggerating symptoms, or performing poorly on unrelated memory or other cognitive tasks are identifiers of malingering and constitute indisputable methods for challenging the veracity of a person's self-reported psychological or somatic symptoms on the MMPI-2 or elsewhere. It is unclear if there are sufficient items within the MMPI-2 item pool to detect the feigning of symptoms of mild brain injury or other somatic complaints. Although robust, the MMPI-2 item pool is finite and additional items (e.g., rare and more diverse somatic symptoms) may be necessary for the precise discriminations the proponents of the FBS claim it can make. Studies are needed to eliminate any items that show bias against women or those with legitimate physical or mental health problems, or any other classes of people who might be harmed by being inappropriately labeled as

malingerers. It is unclear if further research can clarify which, if any, of the myriad suggested cut-off scores that vary by settings and/or medical history and/or psychiatric history and/or radiological findings are the best for identifying malingering and for eliminating the harm that will result when imprecise cutting scores are used and individuals are mislabeled as malingerers and denied necessary treatment or appropriate compensation. And, finally, it is questionable whether psychologists have been adequately prepared to make such confusing discriminations when presented with FBS scores on an individual patient's MMPI-2 profiles. We advise that the prudent and well-informed psychologist avoid using the FBS scale, consistent with our review of the content and research on the scale and with recent court decisions.

## References

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Arbisi, P. A. (2006). Use of the MMPI-2 in personal injury and disability evaluations. In J. N. Butcher (Ed.), *MMPI-2: The practitioner's handbook* (pp. 407–442). Washington, D.C.: American Psychological Association.
- Arbisi, P. A., & Ben-Porath, Y. S. (1997). Characteristics of the MMPI-2 F(p) Scale as a function of diagnosis in an inpatient sample of veterans. *Psychological Assessment*, 9, 102–105.
- Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. (2006). The ability of the MMPI-2 to detect feigned PTSD within the context of compensation seeking. *Psychological Services*, 3, 249–261.
- Arbisi, P. A., & Butcher, J. N. (2004). Failure of the FBS to predict malingering of somatic symptoms: Response to critiques by Greve and Bianchini and Lees-Haley and Fox. *Archives of Clinical Neuropsychology*, 19, 341–345.
- Bagby, R. M., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment*, 7, 84–92.
- Bagby, R. M., Marshall, M. B., Bury, A., Bacchiocchi, J. R., & Miller, L. (2006). Assessing underreporting and over reporting styles on the MMPI-2. In J. N. Butcher (Ed.), *MMPI-2: The practitioner's handbook* (pp. 39–69). Washington, D.C.: American Psychological Association.
- Ben-Porath, Y. S., & Tellegen, A. (2007a). MMPI-2 Fake Bad Scale (FBS). Retrieved December 4, 2007, from [http://www.upress.umn.edu/tests/mmpi2\\_fbs.html](http://www.upress.umn.edu/tests/mmpi2_fbs.html).
- Ben-Porath, Y. S., & Tellegen, A. (2007b). MMPI-2 FBS (Symptom Validity Scale). Retrieved December 4, 2007, from <http://www.pearsonassessments.com/resources/fbs.html>.
- Berry, D. T. R., & Schipper, L. J. (2007). Detection of feigned psychiatric symptoms during forensic neuropsychological examinations. In G. J. Larrabee (Ed.), *Assessment of Malingered Neuropsychological Deficits* (pp. 226–263).
- Berry, D. T. R., Wetter, M. W., Baer, R. A., Youngjohn, J., Gass, C. S., Lamb, D. G., et al. (1995). Over reporting of closed-head injury symptoms on the MMPI-2. *Psychological Assessment*, 7, 517–523.
- Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry*, 153, 7–10.

- Burandt, C. A. (2006). Detecting incomplete effort on the MMPI-2: An examination of the Fake-Bad Scale in electrical injury. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 67(4-B), 2216.
- Bury, A. S., & Bagby, R. M. (2002). The detection of feigned uncoached and coached post-traumatic 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., et al. (2005). NAN position paper: Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, 20, 419–426.
- Butcher, J. N., Arbsi, P. A., Atlas, 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., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for Administration, Scoring, and Interpretation*. Minneapolis, MN: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for Administration, Scoring, and Interpretation (Rev. Ed.)*. Minneapolis, MN: University of Minnesota Press.
- Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S Scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol 10, pp. 25–50). Hillsdale, N. J.: Erlbaum.
- Butcher, J. N., & Williams, C. L. (2000). *Essentials of MMPI-2 and MMPI-A Interpretation* (2nd Ed.). Minneapolis: University of Minnesota Press.
- Charles, T. L., Jr. (2000). Usefulness of the Minnesota Multiphasic Personality Inventory-2 in detection of deception in a personal injury type forensic population. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, 60(10-B), 5221 May.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1973). *An MMPI handbook: Volume I: Clinical interpretation*. Minneapolis: University of Minnesota Press.
- Davidson v. Strawberry Petroleum et al. (2007). Case #05-4320 Hillsborough County, Florida.
- Dearth, C. S., Berry, D. T. R., Vickery, C. D., Vagnini, V. L., Baer, R. E., Orey, S. A., et al. (2005). Detection of feigned head injury symptoms on the MMPI-2 in head injured patients and community controls. *Archives of Clinical Neuropsychology*, 20, 95–110.
- Elhai, J. D., Gold, P. B., Frueh, B. C., & Gold, S. N. (2000). Cross-validation of the MMPI-2 in detecting malingered posttraumatic stress disorder. *Journal of Personality Assessment*, 75, 449–463.
- Fox, D. D. (2005). Distributional characteristics and factor analysis of MMPI-2 F family and Fake Bad Scale. Unpublished data.
- Fox, D. D., Gerson, A., & Lees-Haley, P. R. (1995). Interrelationship of MMPI-2 validity scales in personal injury claims. *Journal of Clinical Psychology*, 51, 42–47.
- Frye v. United States* (1923). 293 F. 1013 (D.C. Cir.).
- Gass, C. (1991). MMPI-2 interpretations in closed head trauma: A correction factor. *Psychological Assessment*, 3, 27–31.
- Gass, C. S. (1996). MMPI-2 variables in attention and memory test performance. *Psychological Assessment*, 8, 135–138.
- Gass, C. S., & Wald, H. S. (1997). MMPI-2 interpretation and closed-head trauma: Cross-validation of a correction factor. *Archives of Clinical Neuropsychology*, 12, 199–205.
- Geisinger, K. F. (2005). The testing industry, ethnic minorities, and individuals with disabilities. In R. P. Philips (Ed.), *Defending standardized testing*. Mahwah, N. J.: Erlbaum.
- 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.
- Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th Ed.). New York: Oxford University Press.
- Graham, J. R., Watts, D., & Timbrook, R. E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment*, 57, 264–277.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd Ed.). Boston: Allyn and Bacon.
- Greiffenstein, M. F., & Baker, W. J. (2001). Comparison of premorbid and postinjury MMPI-2 profiles in late post concussion claimants. *The Clinical Neuropsychologist*, 15, 162–170.
- Greiffenstein, M. F., Baker, W. J., Axelrod, B., Peck, T. A., & Gervais, R. (2004). The Fake Bad Scale and MMPI-2 F-Family in Detection of Implausible Psychological Trauma Claims. *The Clinical Neuropsychologist*, 18, 573–590.
- Greiffenstein, M. F., Baker, W. J., Gola, T., Donders, J., & Miller, L. J. (2002). The FBS in atypical and severe closed head injury litigants. *Journal of Clinical Psychology*, 58, 1591–1600.
- Greiffenstein, M. F., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 Fake Bad Scale in Detection of Noncredible Brain Injury Claims. In K. Boone (Ed.), *Detection of noncredible cognitive performance* (pp. 210–235). New York: Guilford Press.
- Greve, K. W., & Bianchini, K. J. (2004). Response to Butcher et al. (2003). The construct validity of the Lees-Haley Fake-Bad Scale. *Archives of Clinical Neuropsychology*, 19, 337–339.
- Greve, K. W., Bianchini, K. J., Love, J. M., Brenna, A., & Heinley, M. T. (2006). Sensitivity and specificity of the MMPI-2 validity indicators to malingered neurocognitive dysfunction in traumatic brain injury. *The Clinical Neuropsychologist*, 20, 491–512.
- Grossman, L. S., & Wasyliv, O. E. (1988). A psychometric study of stereotypes: Assessment of malingering in a criminal forensic group. *Journal of Personality Assessment*, 52, 549–563.
- Guéz, M., Brännström, R., Nyberg, L., Toolanen, G., & Hildingsson, C. (2005). Neuropsychological functioning and MMPI-2 profiles in chronic neck pain: A comparison of whiplash and non-traumatic groups. *Journal of Clinical and Experimental Neuropsychology*, 27, 151–163.
- Gynther, M. D., & Petzel, T. P. (1967). Differential endorsement of MMPI F Scale items by psychotics and behavior disorders. *Journal of Clinical Psychology*, 23, 185–188.
- Heyward, V. H., & Stolarczyk, L. M. (1996). *Applied body composition assessment*. Champaign, IL: Human Kinetics.
- Hoffman, R. G., Scott, J. G., Emick, M. A., & Adams, R. L. (1999). The MMPI-2 and closed-head injury: Effects of litigation and head injury severity. *Journal of Forensic Neuropsychology*, 1, 3–13.
- Hunt, H. F. (1948). The effects of deliberate deception on MMPI performance. *Journal of Consulting and Clinical Psychology*, 12, 396–402.
- Iverson, G. L., Henrichs, T. F., Barton, E. A., & Allen, S. (2002). Specificity of the MMPI-2 Fake Bad Scale as a marker for personal injury malingering. *Psychological Reports*, 90, 131–136.
- Kay, T., Newman, B., Cavallo, M., & Ezrachi, O. (1992). Toward a neuropsychological model of functional disability after mild traumatic brain injury. *Neuropsychology*, 6, 371–384.
- Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist*, 12, 179–188.
- Larrabee, G. J. (2007). Evaluation of exaggerated health and injury symptomatology. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficit* (pp. 264–286). New York: Oxford University Press.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J.

- Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–372). New York: Oxford University Press.
- Lees-Haley, P. R. (1992). Efficacy of MMPI-2 validity scales and MCMI-II modifier scales for detecting spurious PTSD claims: F, F-K, Fake Bad Scale, ego strength, subtle-obvious subscales, DIS, and DEB. *Journal of Clinical Psychology, 48*, 681–689.
- 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.
- Lees-Haley, P. R., & Fox, D. D. (2004). Commentary on Butcher, Arbisi, Atlis, and McNulty (2003) on the Fake Bad Scale. *Archives of Clinical Neuropsychology, 19*, 333–336.
- Leininger, B. E., Kreutzer, J. S., & Hill, M. R. (1991). Comparison of minor and severe head injury emotional sequelae using the MMPI. *Brain Injury, 5*, 199–205.
- Martens, M., Donders, J., & Millis, S. R. (2001). Evaluation of invalid response sets after traumatic head injury. *Journal of Forensic Neuropsychology, 2*, 1–18.
- McCrae, M. A. (2008). *Mild traumatic brain injury and post concussion syndrome*. New York: Oxford University 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, L. J., & Donders, J. (2001). Subjective symptomatology after traumatic head injury. *Brain Injury, 15*, 297–304.
- Moskowitz, J. L., Lewis, R. J., Ito, M. S., & Ehrmentraut, J. (1999). MMPI-2 profiles of NGRI and civil patients. *Journal of Clinical Psychology, 55*, 659–668.
- 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.
- 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.
- Nicholson, K., & Martelli, M. F. (2007). The effect of compensation status. In G. Young, A. Kane, & K. Nicholson (Eds.), *Causality in psychological injury: Presenting evidence in court* (pp. 411–426). New York: Springer Science & Business Media.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd Ed). New York: McGraw Hill.
- Pearson Assessments (2007). FBS (Symptom Validity Scale) Added to MMPI-2 Standard Scoring Materials. Retrieved December 4, 2007, from <http://pearsonassessments.com/news/pr011107.html>.
- Pope, K. S., Butcher, J. N., & Seelen, J. (2006). *The MMPI, MMPI-2, and MMPI-A in court: A practical guide for expert witnesses and attorneys* (3rd Ed.). Washington, DC: American Psychological Association.
- Post, R. D., & Gasparikova-Krasnec, M. (1979). MMPI validity scales and behavioral disturbance in psychiatric inpatients. *Journal of Personality Assessment, 43*, 155–159.
- Price, D. D. (1999). *Psychological mechanisms of pain and analgesia*. Seattle, WA: IASP Press.
- Price, D. D., & Bushnell, M. C. (2004). *Psychological methods of pain control: Basic science and clinical perspectives*. Seattle, WA: IASP Press.
- Putnam, S. H., & Millis, S. R. (1994). Psychosocial factors in the development and maintenance of chronic somatic and functional symptoms following mild traumatic brain injury. *Advances in Medical Psychotherapy, 7*, 1–22.
- Rissmiller, D. J., Wayslow, A., Madison, H., Hogate, P., Rissmiller, F. R., & Steer, R. A. (1998). Prevalence of malingering in inpatient suicide ideators and attempters. *Crisis, 19*, 62–66.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 160–177.
- Ross, S. R., Putnam, S. H., Gass, C. S., Bailey, D. E., & Adams, K. M. (2002). MMPI-2 indices of psychological disturbance and attention and memory test performance in head injury. *Archives of Clinical Neuropsychology, 18*, 905–906.
- Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*, 213–223.
- Tsushima, W. T., & Tsushima, V. G. (2001). Comparison of the Fake Bad Scale and other MMPI-2 validity scales with personal injury litigants. *Assessment, 8*, 205–212.
- Vandergracht v. Progressive Express et al. (2005). Case no. #02-04552, Hillsborough County, Florida.
- Viglione, V., Muratori, F., Maestro, S., Brunori, E., & Picchi, L. (2006). Denial of symptoms and psychopathology in adolescent anorexia nervosa. *Psychopathology, 39*, 255–260.
- Vitousek, K. M., & Stumpf, R. E. (2005). Difficulties in the assessment of personality traits and disorders in eating-disordered individuals. *Eating Disorders: The Journal of Prevention, 13*, 37–60.
- Wasyliw, O. E., Grossman, L. S., Haywood, T. W., & Cavanaugh, J. L. (1988). The detection of malingering in criminal forensic groups: MMPI validity scales. *Journal of Personality Assessment, 52*, 321–333.
- Williams v. CSX Transportation, Inc. (2007). Case #04-CA-008892. Hillsborough County, Florida.
- Wood, R. L. (2004). Understanding the 'miserable minority': A diathesis-stress paradigm for post-concussive syndrome. *Brain Injury, 18*, 1135–1153.
- 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.
- Youngjohn, J. R., Davis, D., & Wolf, I. (1997). Head injury and the MMPI-2: Paradoxical severity effects and the influence of litigation. *Psychological Assessment, 9*, 177–184.