

The MMPI-2 Symptom Validity Scale (FBS) Is an Empirically Validated Measure of Overreporting in Personal Injury Litigants and Claimants: Reply to Butcher et al. (2008)

Yossef S. Ben-Porath · Kevin W. Greve ·
Kevin J. Bianchini · Paul M. Kaufmann

Received: 3 December 2008 / Accepted: 16 January 2009 / Published online: 13 March 2009
© Springer Science + Business Media, LLC. 2009

Abstract We address issues raised by Butcher et al. (Psychological Injury and the Law 1:191–209, 2008) in their critique of the Minnesota Multiphasic Personality Inventory-2 Symptom Validity Scale (FBS) and show that their analyses and conclusions are based on faulty premises, a misunderstanding of basic concepts in the assessment of overreporting, a selective review of the literature and mischaracterization of the findings they do cite, problematic analyses of a dataset that had already been similarly analyzed, and a flawed analysis of a legal case they discuss. We complement the review of existing research with some new findings that provide further empirical support and clarification of current interpretive recommendations for proper use of the FBS in evaluations of personal injury litigants and claimants.

Butcher et al. (2008) purport to explore potential bias in Minnesota Multiphasic Personality Inventory (MMPI)-2

assessments using the Symptom Validity Scale (FBS)¹. They consider issues related to the item content of the scale, methods used to identify malingering in FBS studies, evidence of the reliability and validity of the scale, and questions about recommended cutoffs for FBS interpretation. Butcher et al. (2008) also seek to present evidence of the false positive rate of the FBS in two samples of individuals they believe to be at risk for mischaracterization as “malingerers” based on scores on the scale. Finally, they discuss the results of a recent Frye hearing in Florida, following which a trial judge excluded testimony based on FBS.

In this rebuttal, we address the issues raised by Butcher et al. (2008) and show that their analyses and conclusions are based on faulty premises, a misunderstanding of basic concepts in the assessment of overreporting, a selective review of the literature and mischaracterization of the findings they do cite, problematic analyses of a dataset (Butcher et al. 2003) that had already been similarly analyzed, and an incomplete analysis of a court case they discuss. Before turning to the more specific problems with Butcher et al.’s (2008) critique, we begin with a discussion and illustration of basic conceptual flaws in their article.

Y. S. Ben-Porath (✉)
Kent State University,
44242 Kent, OH, USA
e-mail: ybenpora@kent.edu

K. W. Greve · K. J. Bianchini
University of New Orleans,
New Orleans, LA, USA

K. W. Greve · K. J. Bianchini
Jefferson Neurobehavioral Group,
Metairie, LA, USA

P. M. Kaufmann
University of Nebraska,
Lincoln, NE, USA

¹ The FBS was originally labeled “Fake Bad” by Lees-Haley et al. (1991). However, shortly after it was added to the MMPI-2 standard set of validity scales, its name was changed to “Symptom Validity”, to address concerns that the original label, although in keeping with a widely used nomenclature might be viewed as prejudicial in psycholegal assessments. Butcher et al.’s (2008) reference list includes a citation of the official announcement of the addition of the FBS to the MMPI-2 materials (Ben-Porath and Tellegen 2007) that includes the corrected full name of the scale.

Conceptual Problems with the Use of the Term “Malingering”

A central theme of Butcher et al.’s critique of the FBS is their assertion that many individuals with real problems will be diagnosed as “malingering” on the basis of a high FBS score that actually reflects an accurate report of their symptoms. This is a straw man argument that disregards an extensive literature on the diagnosis of malingering as well as specific recommendations for FBS interpretation. Research on malingering detection and diagnosis has progressed considerably in the last 15 years. The majority of this research has relied on clear and objective operationalizations of malingering in clinical samples. These methods were formalized in the diagnostic systems for Malingered Neurocognitive Dysfunction (MND; Slick et al. 1999) and Malingered Pain-Related Disability (MPRD; Bianchini et al. 2005). The essential conceptual elements of these systems were summarized by Larrabee et al. (2007):

Both the Slick criteria for MND, and the Bianchini et al. (2005) criteria for MPRD do not require determination by the clinician of whether or not a specific/single behavior is indicative of intentional exaggeration. Instead, the ultimate determination of intent is dealt with in a more comprehensive manner by considering multiple, highly improbable events as indicative of intent. This is a clear departure from past attempts to define malingering which have frequently placed the clinician in the position of inferring intention from a single event. Rather, it is not necessary to rely on single events that unequivocally demonstrate intent because the MND and MPRD criteria are based on behaviors and symptom report that are atypical for and not representative of expected clinical findings in legitimate, unequivocal neurological, psychiatric or developmental disorders. Thus intent is inferred as a result of the combined improbability of events rather than relying on a single definitive indication of intent. (p. 338).

A critical feature of these diagnostic systems is that they rely on “behaviors *and* symptom report that are atypical for and not representative of expected clinical findings in legitimate, unequivocal neurological, psychiatric or developmental disorders” (p. 338; Larrabee et al. 2007; emphasis added). In other words, a positive finding is one rarely (*not never*) seen in patients who are not malingering. Ultimately, regardless of the false positive error rate of any single indicator of response bias, malingering detection techniques are not perfect and should not be used in isolation for the clinical diagnosis of malingering. As has been noted in many articles, manuals, and book chapters and systematized in the Slick et al. (1999) and Bianchini et al. (2005) criteria, a formal

diagnosis of malingering should be based on the integration of diverse sources of information.

This means that in the absence of other evidence of malingering, a positive FBS score alone is insufficient for its diagnosis. Additionally, a diagnosis of malingering cannot be made in the absence of some external incentive. In keeping with this contemporary conceptualization of malingering, when offering recommendations for current use of the FBS, Greiffenstein et al. (2007) indicated explicitly: “Never use FBS alone; combine FBS score with behavior observations and other validity test indicators” (p. 229).

In addition to failing to properly consider the role FBS and similar scales are designed to play in the assessment of malingering or response bias, Butcher et al. (2008) also apply a double standard in their analysis. To illustrate, consider the following excerpt from Butcher and Williams’ (2000, p. 45) recommendations for interpreting scores on their preferred MMPI-2 overreporting indicator, the F scale:

T 65–80: Likely a valid profile, but some symptom exaggeration is possible

T 81–90: Borderline validity; suggests possibly confused and disoriented pattern; likely exaggeration of complaints; use of symptoms to gain services, sympathy, etc.

T 91–99: High-ranging profiles, which should be interpreted very cautiously

T 100–109: Probably invalid, but some profiles if inpatient psychiatric patients and incarcerate felons who have recently been admitted can be interpreted up to 109 if VRIN is in the valid range.

In the sample used by Butcher et al. (2003, 2008), nearly 60% of the veterans score are 65T or higher on F (25.9% score 65–80 and 33.7% score at or above 80T). Applying Butcher et al.’s (2008) logic to Butcher and Williams’ (2000) recommendations would indicate that nearly six out of ten of these veterans are at least “possibly exaggerating their symptoms” and a third are “likely exaggerating their complaints to gain services, sympathy, etc”. The only way to avoid reaching such an implausible inference is to realize that an elevated score on *any* MMPI-2 validity scale is *not* synonymous with malingering and to recognize the need to adjust cutoffs depending upon the population being assessed. These are two basic elements of FBS interpretation that Butcher et al. (2008) ignore (in the case of not inferring malingering on the basis of single scale scores) or find problematic (as discussed later in reference to their comments on adjusting interpretive cutoffs).

Butcher et al.’s (2008) mischaracterization of the meaning of a “positive” FBS result carries over to their discussion of base rates, in which the rate of positive FBS

findings is equated with the base rate of diagnosed malingering. Specifically, Butcher et al. (2008) mischaracterize data on the base rate of malingering, focusing on three studies, two of which have nothing to do with the prevalence of malingering. At the same time, they ignore the extensive literature on the prevalence of malingering in various patient types and medicolegal contexts. Their passage on this topic is confusing and wrong:

... 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. (p. 202).

This paragraph appears to reflect a misunderstanding of diagnostic statistics. In the case of Larrabee (1998), the only cases examined were ones that “showed objective evidence of cognitive malingering on symptom validity testing” (p. 181). Thus, “100%” is not a “base rate” and it is misrepresented as such by Butcher et al. Regarding the Greiffenstein et al. (2004) study, their tables (Tables 4 and 5) report predictive power for a range of hypothetical base rates. Base rate (or its synonyms, prevalence, and pretest odds) information is necessary for the computation of predictive power and should be calculated using prevalence estimates that are reasonable for the population to which the test will be applied. Greve and Bianchini (2004a, b) advised reporting predictive power for a range of hypothetical base rates in studies of malingering classification accuracy: “*Predictive Power* (especially +*PP* [positive predictive power]) associated with these cut-off's for a range of likely base-rates of malingering (e.g., 10 to 50 at 10 increments) should then follow” (p. 536). The rates used by Greiffenstein et al. (2004) are consistent with the range reported by Mittenberg et al. (2002).

The survey by Sharland and Gfeller (2007) is but one study that provides base rate estimates of malingering ranging from 1% to 20%. However, a more thorough review of studies relevant to the question of malingering prevalence shows that a large number of Americans (about 40%) believe that purposeful misrepresentation of claims in the compensation system is acceptable (Public Attitude Monitor 1992, 1993). Covert video surveillance demonstrated evidence of malingering in 20% of patients with incentive who were undergoing pain treatment (Kay and Morris-Jones 1998). In patients with pain complaints, rates of malingering may range from 20% to approaching 40% (Mittenberg et al. 2002). Mittenberg et al. (2002) estimated

the base rate of malingering in traumatic brain injury (TBI) to be between 30% and 40% based on survey of board-certified clinical neuropsychologists. Larrabee (2005) reviewed a number of studies of malingering test performance and found a similar rate. Bianchini et al. (2006) reported comparable rates of failure on malingering tests and other validity indicators and diagnosable malingering in TBI and found that rates varied with the magnitude of incentive. Ardolf et al. (2007) and Chafetz (2008) reported base rates of 50% or more in a criminal forensic settings and Social Security Disability evaluations, respectively. Rates in patients claiming injuries due to toxic exposures were in a similar range (Greve et al. 2006a, b). Overall, these studies suggest that the rate of malingering likely ranges from 20% to 50% across a wide range of clinical conditions and medicolegal contexts. As will be illustrated later, when considered in the context of actual data on malingering base rates, the rates of positive FBS findings in cases where there is potential financial compensation are not unreasonable.

Misleading Descriptions and Appraisals of Other MMPI-2 Scales

Throughout their critique, Butcher et al. (2008) seek to draw a distinction between their favored MMPI-2 validity scales (F in particular, but also L, K, and S) and the shortcomings they perceive in FBS. Their examples and illustrations include an erroneous conceptual foundation of the F scale, a lack of awareness of the nature of the “norms” for the original L and F scales of the MMPI, a one-sided analysis of item overlap between FBS and substantive MMPI-2 scales, and a double standard (already discussed and illustrated further here) regarding “high stakes” psychological measures.

An Erroneous Conceptual Foundation for the F Scale

In describing the advantages of their preferred method for assessing overreporting with the MMPI-2, the original MMPI F scale, Butcher et al. (2008) state:

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). (p. 198).

As just cited, Butcher and Williams (2000) do indeed similarly describe the development of the F scale:

Hathaway and McKinley (1942) considered symptom exaggeration or faking an important response tendency to detect in self-report assessment. They developed a simple yet highly effective means of detecting the tendency to claim an inordinate number of psychological symptoms or to exaggerate one's adjustment problems. The idea underlying the F, or Infrequency, scale was that individuals who are attempting to claim psychological adjustment problems that they do not have will actually go to extremes and endorse symptoms from broad and inconsistent problem areas... Hathaway and McKinley conducted an item analysis and selected items that were infrequently endorsed in the normal adult sample... [t]he authors assumed that an individual who subscribed (sic) to a large number of these rarely endorsed symptoms was claiming too many problems." (p. 44).

This description of the rationale for developing the F scale, with a citation of the original MMPI manual (Hathaway and McKinley 1942) as its source, is incorrect. Hathaway and McKinley (1942) never mentioned overreporting in the original MMPI manual (let alone as the intended use for the F scale). Rather, they stated: "If the F score is high, the scales are likely to be invalid either because the subject was careless or unable to comprehend the items, or because of extensive errors in entering the items on the record sheet. *A high F score has no other known interpretation.*" (p. 9, emphasis added).

The F scale was developed to identify MMPI protocols that were invalid due to carelessness or incapacity to respond appropriately on the part of the test-taker or errors in converting item responses (which were generated by having the test-takers sort 550 statements printed individually on index cards into three groups: True, False, and Cannot Say) into scale scores. Meehl and Hathaway (1946) indicated that "it would be desirable to develop a scale for detecting... tendencies to put oneself in a bad light when answering a personality inventory... *The F scale of the MMPI was not originally developed with this in mind*, but subsequent evidence showed that it could be used in this way." (p. 534, emphasis added). Thus, contrary to Butcher et al.'s (2008) assertion, the F scale and the infrequent item response approach underlying it were not developed to assess symptom overreporting, but were later found to be effective in detecting some forms of biased responding. In summary, Butcher et al.'s basic notion that infrequent response measurement is superior to other techniques for identifying biased responding is predicated on an erroneous belief about the conceptual foundation of the F scale.

Norms and Cutoffs

Butcher et al. (2008) criticize FBS because interpretive recommendations for the scale are based on raw scores rather than standardized T scores. They state:

The development of norms is essential for the interpretation of scores on high stakes assessments like the FBS... All of the MMPI-2 validity scales are interpreted using standardized T scores... Interpretation of the FBS has not followed this traditional approach relying, instead, on various raw score cutoffs..." (p. 200)

We agree that it is preferable to convert raw scores to standard scores, primarily for the purpose of comparing and integrating scores across scales. Lees-Haley et al. (1991) did not have access to the MMPI-2 normative sample, and therefore, their initial recommendations and subsequent modifications were expressed in raw scores. A forthcoming MMPI-2 test monograph will provide T score conversion tables and interpretive recommendations for FBS expressed in T scores (Ben-Porath et al. 2009).

However, Butcher et al.'s (2008) assertion that all other validity scale interpretation is based on standardized T scores is incongruous with years of MMPI practice. For the first 47 years of their existence, "T scores" on L and F were simply arbitrary values affixed to raw scores (and hence, in effect, the same as raw scores). For example, Hathaway and McKinley (1942) explained that "since... it is not possible to assign T scores [to F] in the usual manner, the [T score conversion] table (Table X) has been made up on the basis of experience." (p. 12). Nevertheless, these scales were used routinely across a wide range of settings including in high stakes forensic assessments. Nine years later, they elaborated: "The T scores given for the ?, L, and F scales were arbitrarily assigned and do not derive from any formula. Clinical experience has shown that the T scores assigned for the L and F scales were not appropriately chosen... In order to make more accurate interpretations of these two scales, therefore, any decisions involving their use should henceforth be made in terms of raw scores rather than T scores." (Hathaway and McKinley 1951, p. 12). In fact, standard T scores were not derived for these scales until 1989, based on the MMPI-2 normative sample (Butcher et al. 1989). Therefore, although Butcher et al. (2008) are correct in asserting that all other MMPI-2 validity scales are interpreted based on T scores, they ignore the fact that throughout the nearly 50-year history of the original MMPI interpretation of F and L was essentially based on raw scores.

Butcher et al. (2008) also express concern that "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.” (p. 201). They seek to illustrate this problem by citing various cutoffs proposed for FBS interpretation. Their concerns over a “variable yardstick” are also incongruous with the history of their preferred overreporting indicator. As has been the case with FBS, interpretive recommendations for F evolved with the accumulation of clinical experience and research, and they remain contingent upon the facts and context of the case.

The original cutoff recommended by Hathaway and McKinley (1942, 1943) for identifying invalid MMPI protocols was T score 70. In the 1951 edition of the manual, they raised this cutoff to 80. No interpretive recommendations were offered for F in the 1967 edition of the manual. In the first edition of the MMPI-2 manual, Butcher et al. (1989) stated that F T scores in the 71–90 range indicate questionable validity (possibly due to malingering) whereas T score values of 91 and above indicate that the protocol is probably invalid (with no mention of malingering as a possible reason). Finally, in the current revised edition of the MMPI-2 manual, Butcher et al. (2001) provide different cutoffs for interpreting scores on F for nonclinical, outpatient clinical, and inpatient clinical settings. Of note, the interpretive recommendations for F provided by Butcher and Williams (2000) and listed earlier are not consistent with those of either the 1989 or 2001 editions of the MMPI-2 manual. Butcher et al.’s (2008) reservations about the lack of consensus regarding recommended cutoffs for FBS are puzzling when considered in the context of the history of—and current recommendations for their preferred scale—F.

Moreover, the search for an “optimal cutting” score that works in all cases is misguided (Gallop et al. 2003). Regarding the MMPI-2 and malingering, Greve et al. (2006a, b) pointed out that “for these findings to be clinically useful it is not necessary to identify a ‘best’ or ‘recommended’ cut point” (p. 509). There is variability in the specific scores examined in particular studies. There are also studies that report classification accuracy data for a range of FBS scores (e.g., Bianchini et al. 2008; Greiffenstein et al. 2004; Greve et al. 2006a, b; Larrabee 2003; Ross et al. 2004). As research has progressed, the FBS score range considered to be consistent with malingering has risen. Thus,

to a large extent, the changing FBS cut scores and more subtle interpretations reflect advances in FBS research and should be lauded rather than criticized.

Concerns About Item Overlap

Under the heading of “Item Bias in the FBS”, Butcher et al. (2008) observe: “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.” (p. 193). In a previous critique, Butcher et al. (2003) commented that “the FBS has a considerable item overlap (almost 1/3 of the items) with the three scales that measure health concerns or physical symptoms—two clinical scales (Hs and Hy) and one Content Scale (HEA).” (p. 479)

Absent from Butcher et al.’s current or prior analysis of item overlap is any consideration of how their preferred overreporting indicator, F, compares with FBS in this regard. Given its sensitivity to overreporting of severe psychopathology, a relevant content domain for examining overlap between the F scale items and the substantive scales would be measures of thought dysfunction (i.e., clinical scales 6 and 8 and the content scale Bizarre Mentation). Of the 60 F items, 24 (40%) appear on one or more of these scales.

To bolster their claim that FBS variance is hopelessly confounded with substantive variance, Butcher et al. (2008) observe that “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.” (p. 197). Here too, Butcher et al. fail to provide a context for evaluating these findings. Toward this end, we conducted additional analyses with the sample of psychiatric inpatient male veterans analyzed by Butcher et al. (2003, 2008). In Table 1, we report correlations between the four

Table 1 Correlations between MMPI-2 over-reporting indicators and clinical scales

Overreporting indicators	Clinical scales							
	Hs	D	Hy	Pd	Pa	Pt	Sc	Ma
F	0.52	0.43	0.20	0.66	0.68	0.70	0.85	0.52
Fb	0.54	0.51	0.24	0.63	0.63	0.77	0.85	0.44
Fp	0.20	0.05	−0.03	0.22	0.36	0.26	0.45	0.38
FBS	0.79	0.69	0.75	0.36	0.49	0.60	0.55	0.05

MMPI-2 overreporting indicators and the eight original clinical scales of the MMPI-2. As reiterated by Butcher et al. (2008), scores on FBS are substantially correlated with scores on clinical scales 1 and 3 (0.79 and 0.75, respectively). However, scores on F and Fb are even more highly correlated with scores on clinical scale 8 (0.85 for both validity scales) and comparably correlated with scores on scale 7 (0.70 and 0.77, respectively).

Applying Butcher et al.'s (2008) logic, one would conclude that in light of the item overlap between F and substantive MMPI-2 measures of thought dysfunction and the correlation between this scale and scale 8, variance on F is hopelessly confounded with genuine psychopathology and MMPI-2 users are likely to accuse genuine psychiatric inpatients in particular that they are overreporting. Or, if one were to fully incorporate Butcher et al.'s equation of an elevated score on a validity scale with malingering that these patients are malingering. Recall that such an application of Butcher and Williams (2000) interpretive recommendations for F would result in the following statement "likely exaggeration of complaints; use of symptoms to gain services, sympathy, etc." in reference to one third of these veterans who score 80T or above on F.

Recognizing that individuals experiencing severe psychopathology are likely to produce significantly elevated scores on F, the authors of the current edition of the MMPI-2 manual indicate the need to apply higher cutoffs when interpreting scores on this scale in psychiatric inpatients, as well as the need to consult scores on other scales (i.e., Fp) before reaching any inferences about overreporting. This is consistent with Greiffenstein et al.'s (2007) identification of "moderators" (e.g., medical history) requiring the use of higher cutoffs in FBS score interpretation and their recommendations that scores on all the MMPI-2 overreporting indicators be considered.

Here again, Butcher et al. (2008) set up unrealistic expectations and selectively apply them to FBS. The measurement of any form of exaggeration is likely to include some symptoms of the actual clinical phenomena. Exaggeration is indicated when individuals endorse a pattern or quantity of symptoms that is inconsistent with those presented by typical patients.

Concerns About "High Stakes" Measures

Butcher et al. (2008) state that "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)." (p. 193). We have already discussed the fallacy of equating scores on a single MMPI-2 scale with malingering. Based upon this faulty

premise, these authors state that scores on the scale may deprive individuals of their due compensation, and, therefore, the measure is worthy of particular concern and scrutiny. We agree that like any other measure designed to inform important decisions about individuals, the validity of FBS for its purported applications should be the subject of careful empirical examination. In this context, we note that research on the FBS (summarized later) provides sound empirical support for the scale.

Butcher et al.'s (2008) concern about important decisions informed by FBS reflects another double standard in their analysis. Elsewhere in their critique, they question the inclusion of items reflecting underreporting on the 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. (p. 194)

Butcher et al. (2008) characterize S as a well-validated measure of defensiveness designed to identify such responding in child custody cases and employment screening, two types of evaluation where the stakes are no less high than in personal injury assessments. Indeed, in his interpretive reports for child custody evaluations (Butcher 1998) and for personnel screening (Butcher 2001), the lead critic of the FBS relies on the S scale (as well as subscales that have never been incorporated in the MMPI-2 manual) to characterize child custody litigants and job seekers as defensive. In contrast with the abundant empirical literature on the validity of FBS as a measure of overreporting (reviewed later), no study published to date has examined the validity and utility of the S scale as a measure of underreporting in either custody litigants or individuals undergoing preemployment evaluations.

In providing a rationale for developing the S scale, Butcher and Han (1995) observed: "One problem with the K scale is that it was not developed for use with non-inpatient psychiatric samples (e.g., nonclinical groups such as family custody cases or applicants for employment who have a clear motivation to assert extremely good adjustment in order to present a favorable picture of themselves usually have extreme K scores). There is no

research to apply K in this context, or even to ensure that any K correction should be made” (p. 26). The same can still be said of S, and Butcher et al.’s (2008) concerns about use of FBS in high stakes evaluations are thus puzzling when considered in light of the fact that in his reports Butcher routinely interprets scores on S, as well as K-corrected clinical scale scores of family custody cases and job applicants.

As a final example of how concerns expressed by Butcher et al. (2008) about FBS are inconsistent with their recommendations and practices with other high stakes MMPI-2 scales in offering interpretive recommendations for the *MacAndrew Alcoholism Scale* (MAC), Butcher and Perry (2008) indicate:

Initially MacAndrew recommended a cutoff of 24 as indicative of alcohol abuse problems. This cutoff is probably too low because it is less than one standard deviation above the mean for the original Minnesota normals. A more conservative cutoff is therefore recommended. A general rule of thumb for interpreting the MAC scale is as follows:

1. For males, a raw score of 26 to 28 suggests that alcohol or drug abuse are possible; a raw score of 29 to 31 suggests that alcohol or drug abuse problems are likely; a raw score of 32 or more suggests that alcohol or drug abuse problems are highly probable.
2. For females, a raw score of 23 to 25 suggests that alcohol or drug abuse are possible; a raw score of 26 to 29 suggests that alcohol or drug abuse problems are likely; a raw score of 30 or more suggests that alcohol or drug abuse problems are highly probable. (p. 74)

These recommendations are similar to just about every element of the FBS interpretive recommendations that Butcher et al. (2008) criticize. A scale developer recommended initial raw score cutoffs that were later found to be too low; subsequent authors recommended higher and different raw score cutoffs for men and women; Butcher and Perry (2008) continue to recommend raw score cutoffs rather than T scores, and their recommended cutoffs are substantially at odds with the interpretive recommendations for this scale in the test manual (Butcher et al. 2001), which (appropriately) are far more conservative, indicating that the most one could infer about a test-taker who scores high on this scale is the possibility of substance abuse, which needs to be corroborated by extratest data.

To summarize, in this section, we scrutinized Butcher et al.’s (2008) criticisms of the FBS in light of their views, recommendations, and practices regarding other MMPI-2 scales. We found that their examples and illustrations

include an erroneous conceptual foundation of the F scale, ignore interpretive practices with the original L and F scales of the MMPI, advance a misleading analysis of item overlap between FBS and substantive MMPI-2 scales that fails to consider the implications of similar overlap between their favored scale, F, and substantive measures of thought dysfunction, and reflect a double standard regarding “high stakes” psychological measures and assessments. With this context as a backdrop, we turn next to their problematic analysis of the literature on the validity of FBS.

A Distorted Review of the Literature

Butcher et al. (2008) present a review of the scientific literature regarding FBS and recommend strongly against its clinical and forensic use: “we advise that the prudent and well-informed psychologist avoid using the FBS scale” (p. 207). In this section, we demonstrate that this conclusion is based on a literature review that mischaracterizes the existing data on FBS.

Butcher et al.’s (2008) overall argument against the diagnostic accuracy of the FBS involves a faulty differential prevalence comparison. First, as mentioned above, they cite a malingering base rate that is lower than current estimates, based on a selective review of the literature on base rates. Next, they report a high rate of “FBS-positives” based on low cutoffs that have long been discarded in favor of more conservative ones. Then, based on this distorted comparison (artificially low malingering rate versus artificially elevated FBS hit rate), they argue for a problem with false positives on the FBS. This is the central basis for the Butcher et al. (2008) warning that “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 malingerers has been elevated” (p. 206). However, published data using appropriate malingering base rates and FBS cutoffs described below demonstrate that there is not a problem with FBS false positives.

In the following, we demonstrate that a fair and complete review of the literature supports the interpretive guidelines of Greiffenstein et al. (2007), which were recommended to FBS users by Ben-Porath and Tellegen (2007) when the scale was added to the test materials. This review will address misrepresentations regarding FBS cutoffs and their interpretation, claims that FBS is biased against patients with genuine injury, illness, or disability and does not reliably differentiate between malingerers and nonmalingerers, and the specific claim of gender bias associated with the scale. We complement the review of existing research with some new findings that provide further empirical

support and clarification of current interpretive recommendations for the scale.

Concerns About “False Positives”

We have already discussed Butcher et al.’s faulty equating of a “positive” finding on FBS with malingering. From our perspective, appropriate interpretation of an FBS score that exceeds a recommended cutoff is as a signal about the *possibility* of symptom overreporting. This possibility needs to be considered in light of other test results, available background information and records, behavioral observations, and, of course, the context of the evaluation, including whether an incentive exists to malingering. Although their equating of a “positive” finding with malingering is wrong and does not reflect recommended use of the scale, it is important to ask whether injury alone or genuine disability can produce FBS scores that exceed recommended cutoffs for identifying overreported symptoms. Butcher et al.’s assertion that scores on the scale are confounded with disability is contradicted by the existing literature.

To address questions about possible confounds of elevated scores on FBS, Greiffenstein et al. (2007) compiled data on the rates of positive findings for a meaningful range of cut scores in clinical cases without known incentive. They report these data and the composite specificity data in a single table (p. 222; Table 10.3). In the following, we examine some of the same studies reviewed by Greiffenstein et al. (2007) and have included data from more recently published papers as well as some as yet unpublished data sets. We appreciate the generosity of the researchers acknowledged in our author notes who took the time to provide us with their data for this analysis.

Greiffenstein et al. (2007) presented the percentage of positive findings for a range of FBS scores in 1,052 cases from diverse samples of neurological, medical, psychiatric, and other patients known to be without external incentive to exaggerate. Of those cases, 8.5% scored in the range from

23 to 28 (inclusive) and only 1.2% scored greater than 28. The results from a combined sample of 77 patients with moderate–severe TBI (Greve et al. 2006a, b; Ross et al. 2004) without external incentive were the same (see Table 2 for details). Similar findings have been reported in patients with objectively documented epilepsy (53 patients with medically intractable epilepsy; Nelson et al. 2006; exact scores provided courtesy of Nat Nelson). Barr’s (2005) epilepsy sample had twice as many scores in the 23–28 range (18%) but still had fewer than 5% with scores greater than 28. For the data from the combined group, see Table 2. These results are almost identical to those for a no-incentive psychiatric outpatient sample (Greve et al. 2006a, b; Tsushima and Tsushima 2001). In a combined chronic pain sample without incentive (Bianchini et al. 2008; Meyers et al. 2002), the percentage of patients scoring in the 23–28 range was 23.9% while 1.5% scored above 28. Overall, these data indicate that in patients with genuine injury or illness and who are without incentive to appear disabled, the rate of FBS scores in the 23–28 range is less than 20% (except chronic pain, 23.9%); however, elevations above this level are very rare (generally 3% or less).

FBS scores also tend to be lower in persons with objectively documented injury or illness compared to those whose conditions are associated with minimal, ambiguous, or no objective evidence. Greiffenstein et al. (2007) showed an inverse dose–response relation between the initial severity of TBI (as defined by Glasgow Coma Scale [GCS] score, with noninjured persons ranked lowest and severe TBI ranked highest) and FBS score in 481 TBI cases. That is, the more minor the injury, the higher the FBS score ($R=-0.34$). This relationship was not observed for other MMPI-2 validity scales (F, $R=-0.07$; Fb, $R=-0.08$). For the 191 TBI cases in the Greve et al. (2006a, b) sample who had GCS scores bivariate correlations were similar: FBS, $r=0.43$; F, $r=-0.02$; Fb, $r=0.07$. In this sample, about three times as many patients with mild TBI were positive at each cutoff compared to moderate–severe TBI patients

Table 2 Rate of positive findings at two different FBS score ranges in patients without financial incentive compared to the combined no-incentive sample compiled by Greiffenstein et al. (2007)

Study/source	Diagnosis	N	FBS score range			
			23–28	<i>d</i>	29+	<i>d</i>
Greiffenstein et al. 2007	Mixed clinical	1,052	8.5		1.2	
Greve et al. 2006a, b ^a ; Ross et al. 2004 ^a	Moderate–severe TBI	77	9.1	0.02	0	0.22
Greve et al. 2006a, b ^a ; Tsushima and Tsushima 2001 ^a	Mixed psychiatric	230	13	0.14	1.3	0.01
Nelson et al. 2006; Barr 2005 ^a	Epilepsy	104	15.3	0.21	2.9	0.12
Bianchini et al. 2008; Meyers et al. 2002	Chronic pain	118	23.9	0.43	1.5	0.01

d Cohen’s *d* for the comparison of the Greiffenstein et al. (2007) rate of positive findings at a given FBS score range to the other sample rates for the same FBS score range

^a Sample included in the Greiffenstein et al. (2007) mixed group

without consideration of incentive status (see Table 3). The injury severity effect was also examined in a combined TBI sample using data from Greve et al. (2006a, b), Larrabee (2003), and Ross et al. (2004) in which injury severity was defined on the basis of multiple acute clinical characteristics (e.g., length of loss of consciousness and/or posttraumatic amnesia, radiologic findings) rather than just GCS. Table 3 shows the rate of positive FBS scores continued to be significantly higher in the mild TBI group. Similar findings were seen for patients with confirmed epileptic seizures (Barr 2005; Nelson et al. 2006) compared to those with psychogenic nonepileptic seizures (Barr 2005), with the lowest FBS scores seen in confirmed seizure patients. Greiffenstein et al. (2004) examined patients with symptoms of posttraumatic stress disorder (PTSD) attributed to confirmed major trauma versus trivial stress, again showing higher scores in patients with objectively less significant injury. Thus, when Butcher et al.'s differential prevalence argument is examined with comprehensive consideration of the data, appropriate cutoff ranges, and an empirical understanding of malingering base rates, their argument that FBS elevations in clinical samples indicate problems with specificity is not supported.

Butcher et al. (2008) devote considerable attention to the Greiffenstein et al.'s (2004) study just mentioned. Their criticisms misrepresent the methods, findings, and conclusions of these authors. Greiffenstein et al. (2004) used a known groups design, and the incentive status of their subjects (who were clinical patients) was clearly specified and considered in assigning group membership. These elements (known groups design, known financial incentive status, and the use of clinical patients) provide Greiffenstein et al.'s (2004) analyses considerable advantages over others. Butcher et al. (2008) cite (e.g., Bury and Bagby 2002; Guez et al. 2005) as demonstrating problems with the FBS.

Butcher et al. (2008) criticize the apparent subjectivity of the method used by Greiffenstein et al. (2004) to assign

individuals to either probable or improbable PTSD groups. Greiffenstein et al. (2004) acknowledge that the reader may view the method of group assignment as subjective, but utilize a very well-described methodology that adheres to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition gatekeeper descriptions of stressors as well as an overall organizing principle that relies on dose response, with more severe stressors, being linked to greater likelihood of the condition. Moreover, these authors included only patients for whom external documentation regarding the existence of the stressors was available.

Butcher et al. (2008) describe a sampling bias problem that they assert "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." This is an inaccurate statement because the Greve et al. (2006a, b) study explicitly incorporates litigation status, probability of intentional exaggeration, and brain injury severity and finds very good classification accuracy for FBS. This paper, which directly addresses a shortcoming they perceive in the FBS literature, is ignored by Butcher et al. (2008).

Returning to the apparent brain injury severity effect on FBS, Butcher et al. (2008) suggest that the lower FBS scores in patients with more severe injuries are related to a lack of awareness or insight (anosognosia, anosodiaphoria) sometimes seen with more severe brain pathology particularly involving the right hemisphere and frontal lobes. This interpretation is contradicted by available data. The first question one must ask is why would anosognosia affect FBS and not F or Fb (scales known to be sensitive to emotional distress, of which these patients would presumably be similarly unaware)? In both the Greiffenstein et al. (2007) and Greve et al.'s (2006a, b) TBI data, there was a reverse dose–response relationship between TBI severity and FBS (less severity, higher FBS) but not F or Fb. Moreover, if Butcher et al.'s (2008) hypothesis were

Table 3 The rate of positive FBS scores at two score ranges as a function of the presence of objective evidence of neuropathology

Study/source	Diagnosis	N	FBS score range			
			23–28	<i>d</i>	29+	<i>d</i>
Greve et al. 2006a, b	TBI, GCS 3–12	32	15.6	0.47	3.1	0.78
Greve et al. 2006a, b	TBI, GCS 13–15	135	35.6		28.9	
Greve et al. 2006a, b; Larrabee 2003; Ross et al. 2004	Moderate–severe TBI	206	13.5	0.46	7.3	0.54
Greve et al. 2006a, b	Simple mild TBI	146	32.6		26.7	
Nelson et al. 2006; Barr 2005	Epilepsy	104	15.3	0.57	2.9	0.06
Barr 2005	PNES	75	40		4	
Greiffenstein et al. 2004	Major trauma PTSD	48	31.3	0.21	8.3	0.95
	Improbable PTSD	58	41.4		48.3	

d Cohen's *d* for the comparison of the paired samples of etiologically similar patients at a given FBS score range

correct, one would expect that there to be an association between injury severity and scales L and K, which are sensitive to denial of psychopathology and lack of psychological insight. The correlation of GCS with these two scales in the Greve et al.'s (2006a, b) TBI sample is essentially 0 (L, $r=0.02$; K, $r=-0.02$).

A second point relevant to this question is that the rates of positive FBS findings, particularly at the highest score levels, are just as low in patients without any acquired brain damage at all. For example, 13.0% of a combined no-incentive psychiatric sample ($n=230$; Greve et al. 2006a, b; Tsushima and Tsushima 2001; see Table 2) scored in the 23–28 range and with 1.3% at >28 . Again, the combined no-incentive pain sample described above had somewhat more scores in the 23–28 range but the same rate at >28 (1.5%). Thus, in patients with objective neurological, physical, or psychiatric illness or injury but without incentive, the rates of positive FBS scores using the currently recommended cutoffs are consistently low compared to patients claiming similar injuries but who have minimal or ambiguous pathology.

The role of awareness of deficits can be further addressed by comparing patients with similar injuries/illness but who differ in terms of external financial incentive. The Greve et al.'s (2006a, b) TBI subsample with reported GCS scores was divided into two groups based on GCS (mild, GCS 13–15; moderate–severe, GCS 3–12) and in terms of the presence or absence of financial incentive. Within severity level, the no-incentive and incentive groups did not differ in terms of GCS score. None of the no-incentive moderate–severe cases ($n=13$) scored above 22 on FBS. Because the no-incentive sample was small, it was combined with the Ross et al.'s (2004) sample of 59 no-incentive TBI patients, 56 of whom suffered moderate–severe TBI. The proportion of this combined sample ($n=72$) scoring in the 23–28 range was 4.2% with none above 28. In contrast, 15.8% of those with incentive scored in the 23–28 range and 3.1% scored

>28 . Figure 1 illustrates this relationship. Thus, incentive resulted in a higher rate of positive scores, particularly at the lower cutoff in persons who had suffered objectively moderate–severe TBI.

Butcher et al. (2008) also argue 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. For these reasons, the relatively common research practice of using moderate to severe TBI patients as controls for mild TBI litigants constitutes a serious methodological limitation” (p. 194). While this criticism is fair, again, Butcher et al. (2008) fail to note that Greve et al. (2006a, b) reported data on the sensitivity and specificity of FBS to diagnosed malingering for mild TBI patients separately from moderate–severe TBI patients to specifically address this potential problem. FBS differentiated malingerers from nonmalingerers in both groups. In any case, the proposition is empirically testable. If mild TBI triggers a “significant psychological reaction” to which FBS is sensitive, then mild TBI patients should be equally elevated on FBS regardless of incentive status.

Analysis of the Greve et al. (2006a, b) data set revealed significant main effects for injury severity ($F [1, 280]=12.46, p<0.001, \eta^2=0.04$; mild: mean=24.03, standard deviation (SD)=6.8; moderate–severe: mean=18.53, SD=7.5) and for incentive status ($F [1, 280]=12.46, p<0.001, \eta^2=0.04$; no incentive: mean=15.69, SD=6.9; incentive: mean=22.63, SD=7.3) but no interaction ($F [1, 280]=0.11, p=0.74, \eta^2=0.00$). Examination of the frequency of positive FBS scores demonstrated that regardless of incentive status, approximately 1/3 of all mild TBI patients had scores in the 23–28 range (incentive 32.6%, $n=135$; no incentive=36.0%, $n=11$; Cohen's $d=0.08$). At this score level, the incentive effect was much more powerful for the moderate–severe TBI patients (Cohen's $d=0.40$). At the same time, none of the mild or moderate–severe TBI patients without incentive scored above 28 compared to 28.9% of mild TBI and 3.1% of moderate–severe TBI with incentive. Again, refer to Fig. 1 for a graphical representation of these data. The findings for the 23–28 range are consistent with the idea that some psychosocial process associated with mild TBI and not neurological injury alone influences FBS level. Scores above 28 in patients with TBI are exclusively associated with financial incentive to appear disabled regardless of TBI severity. These data refute Butcher et al.'s (2008) contention that the lower FBS scores in moderate–severe TBI are due to neurocognitive deficits such as anosognosia and anosodiaphoria.

The Greve et al.'s (2006a, b) mild TBI no-incentive sample was small so replication in a larger sample is important. Because of the rarity of no-incentive mild TBI cases, the replication was done in patients with chronic pain. Chronic pain samples present similarly to patients

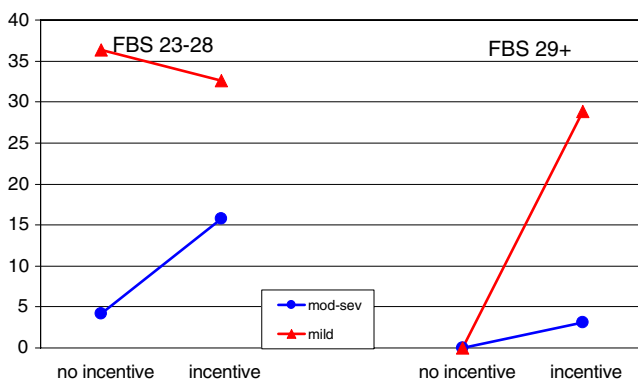


Fig. 1 Percentage of positive FBS at two cutoffs as a function of traumatic brain injury severity (mild, moderate–severe) and incentive status

Table 4 Rates of positive FBS scores at two score ranges as a function of incentive status (present, absent) in traumatic brain injury and chronic pain samples

Study/source	Diagnosis	N	FBS score range			
			23–28	<i>d</i>	29+	<i>d</i>
Greve et al. 2006a, b	Mild TBI, no financial incentive	11	36.4	0.08	0	1.34
Greve et al. 2006a, b	Mild TBI, financial incentive	135	32.6		28.9	
Greve et al. 2008; Meyers and Rohling 2004	Pain, no financial incentive	118	23.9	0.3	1.5	0.86
Bianchini and Greve, unpublished data	Pain, financial incentive	738	37.8		27.5	

d Cohen's *d* for the comparison of the paired samples of etiologically similar patients at a given FBS score range

with persistent postconcussion syndrome (Iverson and McCracken 1997). We compared a combined no-incentive chronic pain sample ($n=118$) derived from Meyers et al. (2002) and Bianchini et al. (2008) with 738 chronic pain patients with financial incentive (Greve and Bianchini, unpublished data)². These data are presented in Table 4. There was no effect for incentive at the 23–28 score range in mild TBI. About a third of all the mild TBI (incentive and no incentive) and pain cases with incentive scored in the 23 to 28 range. Slightly fewer no-incentive pain cases (23.9% compared to 36.4% of no-incentive mild TBI patients) scored in that range. At the higher cutoff, the two patient groups showed nearly identical and large incentive effects. None of the no-incentive mild TBI patients and only 1.5% of the no-incentive pain patients score 29 or higher. In contrast, 28.9% of the mild TBI patients with incentive and 27.5% of the chronic pain patients with incentive scored higher than 29.

These data demonstrate that FBS scores are not elevated beyond 28 by genuine injury or illness alone and that even psychiatric illness *alone* (see Table 2) does not result in

differentially higher scores. Higher FBS scores are associated with the presence of external incentive; this relationship is the strongest in conditions which are either relatively mild or which are associated with minimal objective pathology. Butcher et al.'s (2008) claims notwithstanding, the association between potential financial compensation and poorer outcomes in traumatic brain injury is well-established (Belanger et al. 2005; Binder and Rohling 1996; Binder et al. 1997; Carroll et al. 2004; Tsanadis et al. 2008) as well as increased reports of pain and decreased treatment efficacy in chronic pain (Harris et al. 2005; Rainville et al. 1997; Rohling et al. 1995; Vaccaro et al. 1997).

Butcher et al. (2008) have argued that “a plausible alternative explanation is that many people who have injuries that result in physical and emotional problems legitimately pursue compensation via litigation” (p. 195) and that “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” (p. 200). These arguments are not plausible considering that the incentive effects observed in TBI control for the objective severity of the injury. Greve et al. (2006b) were explicit in addressing this point: “To help address the risk of false positive errors, the tables presented in this paper allow for a careful matching of a given clinical patient with the appropriate comparison group or groups. Such comparisons facilitate interpretation of scores by helping to rule out or rule in alternative explanations for a given score. The make-up of the non-malingering TBI groups is particularly helpful because the groups include persons *with incentive*. That means that the potential stress associated with a workers compensation claim or personal injury litigation is addressed” (emphasis added, pp 505–506).

Bianchini et al. (2006) demonstrated a dose–response association between the magnitude of potential financial compensation and test findings reflective of underperformance or symptom exaggeration. In fact, they showed a doubling of diagnosable malingering (per Slick et al. 1999 criteria) in mild TBI from relatively lower incentive

² Clinical patients were sequential cases seen for psychological pain evaluations at a clinical psychology practice in the southeastern USA from 1998 through 2007. The mean age for the sample was 42.3 years ($SD=9.8$). The sample had completed an average of 11.7 years of education ($SD=2.6$) and were 37.4 months postinjury ($SD=39.4$). The sample was 63.2% male and 66.6% Caucasian (African-American=25.6%; Hispanic=1.9%; other or not indicated=5.9%). These patients were referred by physicians (34.7%), workers compensation companies (47.7%), and attorneys (17.7%). Over half (53.4%) were represented by an attorney at the time of the evaluation. Approximately 98% had external incentive, primarily in the form of workers compensation claims (87.0%) or a personal injury law suit (9.7%). Medical records reviewed as part of these evaluations provided objective medical diagnostic test results, as well as physicians' clinical diagnoses and injury descriptions. Eighty percent of the sample presented with a claim of spinal injury and 91.2% reported subjective neck or back pain. Most (83.4%) had no objective laboratory evidence of spinal pathology (e.g., magnetic resonance imaging/computed tomography scan, electromyography/nerve conduction study). A minority of patients had undergone spine surgery (discectomy/fusion=28.2%; decompression/laminectomy=13.9%). At the time of the interview, the patients rated their pain at 6.7 ($SD=1.9$) out of a maximum of 10.

Table 5 The false positive error rate, sensitivity, and likelihood ratio at different FBS scores ranges in traumatic brain injury and chronic pain

Traumatic Brain Injury ^a						
	not MND			MND		
FBS score	n	FP rate		n	Sens	LR
23-28	36	19.4		52	28.6	1.5
29+	4	2.2		80	44.2	42.1
total n	185			181		
Chronic Pain ^b						
	not MPRD			MPRD		
FBS score	n	FP rate		n	Sens	LR
23-28	41	26.1		6	19	0.7
29+	3	1.9		20	62	32.6
total n	157			32		
Chronic Pain ^c						
	not MPRD			MPRD		
FBS score	n	FP rate		n	Sens	LR
23-28	37	27.8		15	18.5	0.7
29+	0	0		56	69.1	— ^d
total n	133			81		

FP rate false positive error rate (1—specificity), *LR* likelihood ratio (sensitivity/false positive error rate), *MND* Malingered Neurocognitive Dysfunction (Slick et al. 1999), *MPRD* probable and definite Malingered Pain-Related Disability (Bianchini et al. 2005), *n* number of cases scoring in the indicated FBS range, *total n* size of the entire sample, *Sens* sensitivity

^a Combined from the samples of Greve et al. (2006a, b), Larrabee (2003), and Ross et al. (2004). MND group includes probable and definite MND

^b Combined Bianchini et al. (2008) and Meyers and Rohling (2004); MPRD group is only definite MPRD

^c From Bianchini and Greve, unpublished data ($n=476$); MPRD group includes probable and definite MPRD. Patients not meeting criteria for MPRD but in whom malingering could not be reasonably ruled out ($n=262$) were excluded

^d The likelihood ratio cannot be calculated when the false positive error rate is 0 because it requires division by 0. If the false positive error rate were assumed to be 2%, then the LR would be 34.6

(Louisiana state workers compensation claims; 17.7%) to relatively higher potential compensation (Federally based workers compensation claims; 33.3%). No patients without incentive would have met the Slick et al. (1999) psychometric criteria for malingering. The effect of the magnitude of financial incentive was also seen on FBS scores. The proportion of cases with FBS scores greater than 29 steadily increased as a function of workers compensation jurisdiction regardless of injury severity. A worker claiming an injury in a Louisiana jurisdiction was 7.34 times more likely to have an FBS score >29 compared to an injured person with no external incentive. In contrast, a worker injured in a Federal jurisdiction was 12.73 times more likely to have an elevated FBS score compared to a no-incentive control. The odds of a Federal claim being associated with an elevated FBS score were nearly twice that of a state claim. These data indicate that it is not just the presence of incentive and all the stress associated with pursuing an injury claim in the workers compensation or legal arena (e.g., litigation itself, delay in receiving treatment, unemployment, etc.) but the magnitude of

incentive that results in elevated FBS scores. The Bianchini et al.'s (2006) study demonstrates that exaggeration measured by FBS was associated with financial incentive and supports the idea that significantly elevated FBS scores are associated with overreporting rather than either TBI symptoms or simply the psychological effects of a medicolegal context.

Other studies have demonstrated that FBS accurately differentiates between persons who met published criteria for a diagnosis of malingering (i.e., Bianchini et al. 2005; Slick et al. 1999) and patients who were determined not to be malingering. Greve et al. (2006a, b), Larrabee (2003), and Ross et al. (2004) studied FBS in TBI patients while Bianchini et al. (2008) reported FBS data for a known-groups study of patients with chronic pain and Meyers and Rohling (2004) reported data for a sample of pain patients without incentive. All studies reported cumulative frequencies for the full score distribution. We have also examined data for a much larger unpublished sample of pain patients ($n=483$; Bianchini and Greve, unpublished data). Table 5 shows sensitivity, false positive error rates, and likelihood

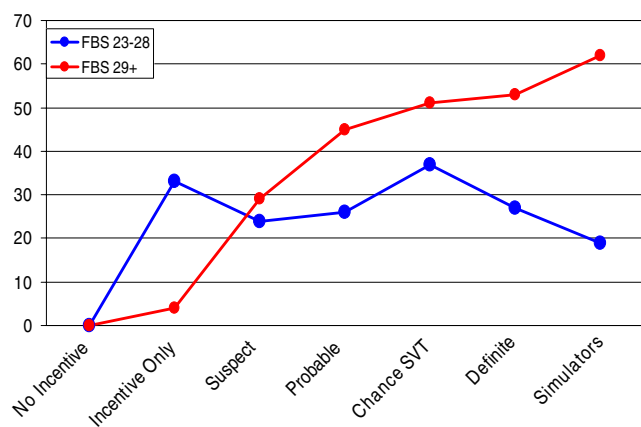


Fig. 2 Relationship between strength of findings reflective of Malingered Neurocognitive Dysfunction and FBS score in traumatic brain injury

ratios for the combined TBI samples, the Bianchini et al. (2008) pain sample, and the combined pain group. Scores in the 23–28 range do not differentiate between groups. This means that these moderate scores should not be treated as an indication of malingering in individuals presenting with pain-related complaints. At the same time, some patients who are malingering produce scores in that range, so scores lower than 29 certainly do not rule out the presence of malingering. However, scores greater than 28 are associated with a small number of false positive errors while detecting between 40% and 50% of diagnosed malingerers.

In fact, a relationship between FBS score and the amount/magnitude of findings consistent with malingering can be demonstrated in both TBI and chronic pain patients. The methodology of Greve et al. (2006a, b), Larrabee (2003), and Ross et al. (2004) allowed the construction of subgroups reflecting a gradation of malingering findings from patients with no incentive to patients with incentive but no evidence of poor effort or underperformance up to those patients who performed significantly worse than chance on a forced-choice symptom validity test (SVT). Simulator data from Bianchini et al. (2008) were also included. As shown in Fig. 2, the proportion of cases with positive findings at both cutoffs increased steadily through the probable MND group and then plateaued. Over 40% of TBI cases who met the Slick et al.'s (1999) criteria for MND scored >28 on FBS. Less than 5% of the patients who were classified as not MND (no-incentive, incentive-only) scored above 28.

Nearly 500 chronic pain patients ($n=476$; Greve and Bianchini, unpublished data) were classified into five groups defined by the amount of evidence of malingering using the Bianchini et al. (2005) criteria for MPRD. The groups were operationalized following the method of Greve et al. (in press) and Greve et al. (2008). MMPI variables were not used for classification. The presence of objective

physical pathology was not associated with group membership. The six groups examined were as follows: (1) no incentive ($n=109$; includes the Meyers and Rohling 2004, sample), (2) negative on all indicators used ($n=95$), (3) a single indeterminate finding present but otherwise negative ($n=29$), (4) multiple indeterminate findings or positive findings but did not meet criteria for probable MPRD ($n=262$), (5) probable MPRD criteria met ($n=44$), (6) definite MPRD criteria met ($n=37$), and (6) simulators ($n=26$; from Bianchini et al. 2008). As can be seen in Fig. 3, a similar dose–response relationship was seen in the pain sample.

FBS Reflects Somatization

Butcher et al. (2008) also argue that other (than intentional overreporting) sources of exaggeration to explain FBS elevations have not been considered: “there are many reasons that lead to symptom exaggeration and preoccupation that fall short of malingering” (p. 194). This is an important criticism of a scale that purports to measure exaggeration of physical symptoms and disabilities. However, here too, Butcher et al. have selectively reviewed the literature, failing to note earlier explicit discussions of this issue. For example, Bianchini et al. (2005) clearly articulated the relevant issues in the context of chronic pain:

The discrepancy between physical findings and physical disability in some patients may be termed “excess disability.” While potentially including some persons whose physical pathology is not visible to current medical diagnostic technology, those patients with excess disability could be reasonably divided in two groups: 1) those whose excess disability is related to *unconscious* psychological factors (i.e., somatization); and, 2) those whose excess disability is the result of *intentional* fabrication or exaggeration... Somatization

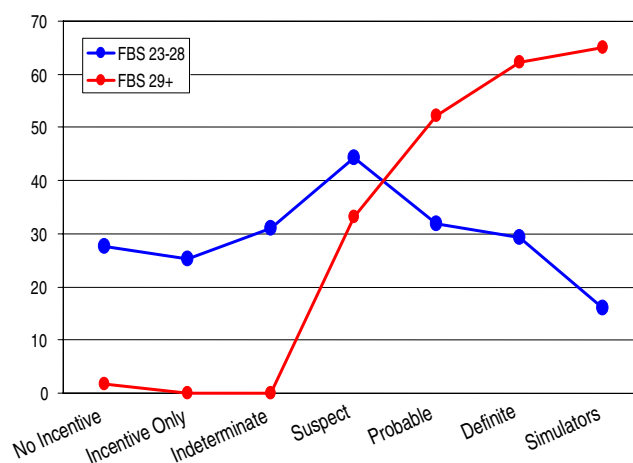


Fig. 3 Relationship between strength of malingering findings and FBS score in chronic pain

is one unconscious psychological process that directly affects pain symptom presentation; it is the characteristic psychological process of somatoform disorders. Conscious mechanisms include intentional attempts to appear impaired to achieve some psychological goal (Factitious Disorder) or to achieve some external incentive (malingering). Psychological mechanisms, including conscious ones, can coexist with documented physical pathology. Similarly, conscious and unconscious psychological mechanisms are not mutually exclusive. Discriminating between unconscious and intentional mechanisms (e.g., hysterical conversion reaction vs. malingering) is one of the central questions that must be addressed. There are existing methods for understanding relevant physical parameters (although not completely) and some aspects of the psychological processes, particularly somatization. What is needed is a system for detecting and diagnosing the intentional or conscious mechanisms in pain. (p. 405)

Both the Slick et al. (1999) and Bianchini et al. (2005) systems conceptually differentiate malingering from somatoform (Bianchini et al. 2005). Butcher et al.'s (2008) critique of FBS as detecting somatization instead of malingering confounds the constructs of malingering and somatization and thus their measurement. Specifically (as discussed earlier), when one uses any symptom exaggeration measure in a given patient population, malingering is reflected in scores that are higher than those typically seen in the nonmalingering clinical presentation including exaggeration of clinical complaints associated with somatization. Scores will demonstrate a stratification of exaggeration etiologies. Lower scores likely reflect neither the influence of somatization nor malingering. Within the next highest score range will be patients whose exaggeration is due to either or both somatization and malingering. However, since scores in this range may reflect exaggeration due to multiple influences, they are not specific enough to be a reliable indication of intentional exaggeration. Thus, there will be false negative cases in this score range. At higher score levels, symptoms are exaggerated to a degree that is very rarely seen except in persons who are known to be malingering. One can thus be confident that scores in this range reflect intentional exaggeration.

Butcher et al.'s (2008) assertion that the FBS provides no information about intent is yet another straw man argument. No scale from a psychological test/symptom survey provides a definitive indication of intent. The best that can be expected is that the scale, at a certain score range, will be a specific indication of membership in a group that is probably malingering. The FBS has been shown to do this in many of the studies cited already.

Moreover, the evidence of physical symptom exaggeration cannot definitively rule out the co-occurrence of somatization with malingering. That is, even when malingering is definitively/clearly detected, somatization cannot be ruled out as a comorbidity. Similarly, the presence of physical pathology does not rule out malingering (Greve et al. 2003; Bianchini et al. 2003; Iverson 2003). It is important for a malingering indicator to be specific for intentional exaggeration of symptoms. However, since somatization and malingering of physical symptoms both involve *exaggeration of physical symptoms*, any indicator that detects malingering is going to also detect exaggeration related to somatization at lower score levels. Only at the higher levels will the score be specific to intentional exaggeration.

Is FBS sensitive to somatization and litigation-related psychosocial stressors? Based on the Bianchini et al. (2008) and Greve et al.'s (2006a, b) data illustrated in Table 5 and Figs. 2 and 3, it is apparent that in scores in the 23–28 range do reflect the influence of factors other than malingering and that scores in this range do not effectively differentiate between malingering and nonmalingering patients with incentive. At the higher cutoff (>28), FBS differentiates between persons intentionally exaggerating their symptoms and those who are not. This stratification and the ability, at the correct cutoffs, to differentiate malingering from somatization is illustrated by the Bianchini et al.'s (2008) sample, where 26% of nonmalingering pain patients scored in the 23 to 28 range compared to 32% of malingers. Scores at this level do not differentiate between the groups. At the same time, a third of the malingering sample would go undetected (are false negative errors) with higher cutoffs. Moreover, the fact that few pain patients without financial incentive score in this range (only 13%) demonstrates the influence of the litigation environment. In contrast, as previous sections have demonstrated, scores greater than 28 in nonmalingers are very rare. In Bianchini et al. (2008), only 2% of the nonmalingering pain patients score higher than 28. In contrast, 62% of definite malingers scored at that level. Very few of the malingers scored below 23 (11%) while 72% of nonmalingers had low FBS scores. Overall, these findings demonstrate the separation of the malingering and nonmalingering FBS distributions which overlap in the 23–28 range.

The research reviewed in this section demonstrates that these very high FBS scores (>28) rarely occur in persons who do not meet published criteria for a diagnosis of malingering. As a result, it is reasonable to conclude that FBS scores greater than 28 are an indication of intentional exaggeration which, in the context of similar findings from other indicators, may lead to a diagnosis of malingering. In summary, elevated levels of symptoms seen likely as a result of somatization are also reflected in elevated scores on the FBS, but at higher scores (29 or greater), the scale

Table 6 FBS means as a function of gender in several published samples

	Male			Female			Diff.	<i>d</i>
	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD		
Normative sample	1,138	11.67	3.81	1462	13.76	4.14	+2.09	+0.53
General clinical	75	15.41	4.62	62	18.19	4.86	+2.78	+0.59
Mild TBI	102	23.03	6.79	58	25.64	6.41	+2.61	+0.39
Pain	301	24.34	6.00	175	25.18	5.91	+0.84	+0.14

The general clinical and mild TBI patients are from Greve et al. (2006a, b)
d Cohen's *d*, *diff* mean female FBS minus mean male FBS, *TBI* traumatic brain injury

specifically detects patients who meet published peer reviewed criteria for malingering. This performance of the FBS in studies using appropriate criterion groups for identifying intentional exaggeration is consistent with what would be expected of a scale that measures physical symptom and disability exaggeration from multiple sources including malingering. Moreover, as reflected in the literature just reviewed, the FBS clearly outperforms the F scale in identifying physical and cognitive symptom exaggeration. Consequently, low correlations between FBS and F that lead Butcher et al. (2008) to question the utility of the former as a validity scale actually provide further evidence of the incremental utility of the FBS.

The Question of Gender Bias

Butcher et al. (2008) contended that because women tend to score higher on FBS, they will be more likely to be diagnosed as malingering. Data on FBS from the MMPI-2 normative sample (as reported in Greene 2000) indicate that women score higher than men by about two raw score points: a medium effect ($d=0.53$; Cohen 1988). An effect of similar magnitude was seen for Greve et al.'s (2006a, b) no-incentive general clinical sample ($d=0.59$) though the absolute values of FBS were slightly higher than was seen in the normative sample. In Greve et al.'s (2006a, b) mild TBI sample, the effect size was in the small range ($d=0.39$) and in a large sample of chronic pain patients, the gender effect was negligible ($d=0.14$). See Table 6 for details of

this analysis. Early recognition of this effect led to adjustment in the recommended raw score cutoffs for FBS (24 men 26 women; Lees-Haley 1992). As described earlier, Butcher and Perry (2008) recommend a similar adjustment (in the opposite direction) for interpreting scores on the MMPI-2 MAC-R scale.

As noted, Butcher et al. (2008) conclude that the higher average score for women translates into a higher rate of "positive" FBS scores. However, as can be seen in Table 7, in the Greve et al.'s (2006a, b) mild TBI sample, the proportion of females scoring above 28 (27.6%) was not significantly different from the males (24.5%; $\chi^2=0.18$, $p=0.69$, $d=0.07$). Moreover, a known-groups analysis demonstrated no gender difference in the false positive error rate at the >28 cutoff. One nonmalingering male (3.2%) and one nonmalingering female (5.0%) earned a score of 29 and none scored any higher. At this cutoff, sensitivity was 37.8% in males and 60% in females. Thus, women who were diagnosed as malingering independent of their FBS scores were more likely to elevate FBS than similarly diagnosed men. This result refutes Butcher et al.'s (2008) argument that FBS biased against women. Rather, it suggests that FBS is more accurate in identifying possible malingering in women with mild TBI compared to similarly injured men.

In the Greve and Bianchini's pain patients overall, there was also no meaningful gender effect on FBS ($d=0.10$; see Table 8). Patients in this data set ($n=476$) were classified as malingering or not malingering using stand-alone and

Table 7 Rates of FBS >28 in males and females presenting with mild traumatic brain injury

	Percentage with FBS ≥ 29			χ^2	<i>d</i>
	Male	Female	F–M difference		
All mild TBI	24.5	27.6	+3.1	0.18	0.07
Mild TBI (not-MND)	3.2	5.0	+1.8	0.10	0.09
Mild TBI (MND)	37.8	60.0	+22.2	2.77	0.45

Criteria for the diagnosis of malingering did not include FBS
d Cohen's *d*, *F–M difference* percentage differential between females and males, *MND* Malingered Neurocognitive Dysfunction, *TBI* traumatic brain injury

Table 8 Rates of FBS >28 in males and females presenting with chronic pain

	Percentage with FBS ≥ 29		F–M difference	χ^2	d
	Male	Female			
All chronic pain	26.2	30.9	+4.7	1.17	0.10
Pain (not MPRD)	14.5	11.3	–3.2	0.44	0.10
Pain (MPRD)	54.5	55.9	+1.4	0.02	0.02

Criteria for the diagnosis of malingering did not include FBS

d Cohen's d , F – M difference percentage differential between females and males, a positive value means more females, *MPRD* Malingered Pain-Related Disability, *TBI* traumatic brain injury

embedded cognitive validity indicators, the MMPI F scale, and evidence from physical examination but not FBS. The false positive error rate in males was 14.5% and in females, it was 11.3%, about 3% less than in males, and the effect size, as noted, was negligible ($d=0.10$). At this cutoff, sensitivity was over 50% for males (54.5%) and females (55.9), again a negligible gender difference ($d=0.02$). Thus, there was not a meaningful gender effect in either the nonmalingering or malingering pain patients at the >28 cutoff. Here too, there is no evidence to support the proposition that FBS is biased against women.

Slightly more than half of both malingering pain groups scored >28 on FBS as did the female malingering mild TBI patients. In contrast, the male malingering mild TBI patients scored that high on FBS at a much lower rate. Again, there were no gender effects at all in the nonmalingering groups and the rates of failure were very low. These relationships are illustrated in Fig. 4. The failure rates in men versus women differ as a function of injury type, raising the intriguing possibility of a gender by malingering status interaction. These findings suggest that the malingering strategy of women with mild TBI and male and female pain patients tends to involve exaggeration of physical symptoms. In contrast, the mild TBI men may rely more heavily on exaggeration of cognitive deficits or psychological symptoms and complain less about physical problems.

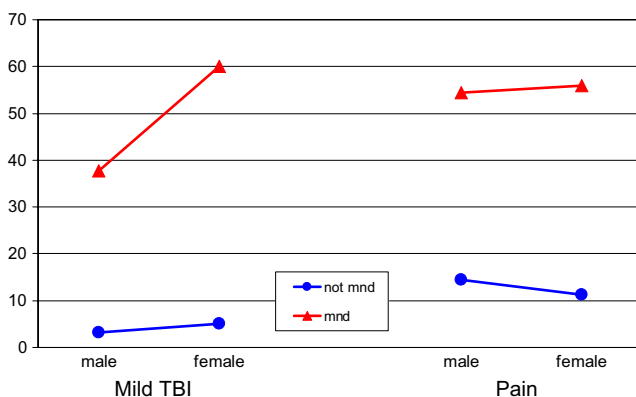


Fig. 4 Percentage of positive FBS at two cutoffs as a function of gender and malingering status in mild TBI and chronic pain

Ultimately, the question is: What is the effect on the base rate of malingering as a function of gender when FBS is included in the diagnosis? Using the classification methodology described above (excluding FBS from the decision making), 25.6% of men and 19.4% of women were classified as malingering, a difference that was statistically nonsignificant ($X^2=2.34$, $p=0.13$, $d=0.15$). When a positive FBS (defined as >28) was included as one of the indicators of malingering, the base rate in men increased to 28.9% (+4.5%) and in women to 25.7% (+7.8%). Neither the gender difference in the base rate ($X^2=0.56$, $d=0.07$) nor the relative increase ($X^2=1.78$; $d=0.14$) was statistically significant or meaningful. In short, while there are gender effects on mean FBS scores, particularly in populations expected to produce low FBS scores (e.g., normals, medical patients without financial incentive), the gender effects disappear at levels of FBS that indicate the possibility of malingering. Women are not disadvantaged by FBS when it is interpreted as recommended. They are not more likely to be diagnosed as malingering if FBS is included among the data considered within a formal malingering diagnostic system.

Butcher et al.'s (2008) Data

To bolster their claims that the FBS has an unacceptably high false positive rate, Butcher et al. (2008) repeat some of the claims made based on analyses reported by Butcher et al. (2003) and present data for a new sample of patients diagnosed with an eating disorder. A false positive error occurs when a test result indicates the presence of a condition when, in fact, the condition is not present (Gallop et al. 2003). The false positive error rate is the complement of specificity (the proportion of cases without the condition who are negative on the test). Butcher et al. (2008) have defined the conditions for this analysis: “Especially of concern are false-positive rates when persons with legitimate head injuries, and resulting somatic symptoms, are mislabeled as malingering” (p. 198). Thus, to determine the rate of false positive errors associated with

FBS requires knowledge of whether subjects in a dataset are malingering.

In a widely criticized paper, Butcher et al. (2003) claimed to demonstrate a high rate of false positive errors in six large patient samples, all but two of which were culled from the MMPI-2 distributor's archival files. However, this claim could not be scientifically tested using their methodology. Butcher et al. (2003) knew little or nothing about their samples including the nature of their injuries or illnesses or the proportion of cases with some sort of financial incentive (the exception being their personal injury sample). More importantly, Butcher et al. (2003) had no idea whether any and, if so, how many of their positive cases were actually malingering. Butcher et al. (2003) noted that their study "is limited by not having a clearly determined 'malingered' and a clearly determined 'nonmalingered' sample on which to verify the classification success" (p. 482). Without these "clearly determined" samples, it is not possible to determine either sensitivity or the false positive error rate (see Greve and Bianchini 2004a, b, for a detailed discussion). All Butcher et al.'s. (2003) data provided is evidence of the base rate of positive findings for given patient types at specific cutoffs. They cannot be used to estimate sensitivity or the false positive rate.

In their present paper, Butcher et al. (2008) offer data on "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" (Abstract). Their VA sample is the same one described in the 2003 paper and Butcher et al. (2008) continue to report rates of positive FBS scores for inappropriately low cutoffs (the highest was >24). Conspicuously absent is any information about higher cutoffs. The second sample of hospitalized eating disorder patients is new. The authors comment that "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" (p. 203).

While the claims of false positive errors in the VA sample suffer from the fatal flaw inherent in not knowing the patients' actual malingering status, the claims related to the eating disorders sample are flawed for a different reason. Butcher et al. (2008) state that Greiffenstein et al. (2007) indicate that 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)" (p. 204; Butcher et al. 2008). Butcher et al. then go on to state that "8% of this eating disorder sample that we studied would be classified as malingerers even using this cut-off score of 'greatest confidence'" (p. 204). They disregard the influence of financial incen-

tive, regardless of malingering status, and do not even report incentive status even though a high proportion of eating disorders patients may have a disability claim and therefore financial incentive (Su and Birmingham 2003).

In presenting the argument about the eating disorder patients, Butcher et al. (2008) again inappropriately equate an elevated FBS score with a diagnosis of malingering in the absence of other evidence of symptom exaggeration or cognitive underperformance. Greiffenstein et al. (2007) state unambiguously: "*General prohibitions.* Never use the FBS alone; combine FBS score with behavior observations and other validity test indicators... positive FBS score does not automatically rule out the coexistence of genuine problems, but it does indicate magnification of problems in such cases." (p. 229). If the eating disorder patients have been accurately characterized, then none of them would receive a diagnosis of malingering no matter how high their FBS score (see Bianchini et al. 2005; Slick et al. 1999).

Butcher et al.'s Critique of the Decision to Add the FBS to the Standard MMPI-2 Scales

In a section of their critique titled "Evaluation of the Publisher's and Distributor's Statements on the FBS", Butcher et al. (2008) imply a connection between "a solicited letter from the developer of the FBS endorsing the MMPI-2-RF", a new version of the MMPI-2 then under development, and the initiation of a review process that led to the addition of the FBS to the standard MMPI-2 scoring materials. This insinuation is false. In fact, the developer of the FBS was one of several psychologists asked to review a preliminary set of scales for the MMPI-2-RF in order to recommend targets for further scale development. No endorsements were solicited. The decision to review the FBS for possible addition to the MMPI-2 standard materials was unrelated to the MMPI-2-RF development process and such innuendo of a quid pro quo has no place in a legitimate scholarly discussion of the scientific merits of the FBS. As cited in the next section, a similar attempt to imply such a connection by the first author of the Butcher et al. (2008) critique and an attorney with whom he frequently collaborates in her advocacy for plaintiffs in personal injury litigation was rebuffed by the judge in the Frye hearing case (*Williams v. CSX Transportation, Inc.* 2007) discussed in their critique.

Butcher et al. (2008) then go on to discuss and quote from reviews obtained by the MMPI-2 publisher in the process of determining whether to add the FBS to the standard test materials. They explain that these documents were provided under the *Minnesota Data Practices Act*, but neglect to mention that they were provided to the same plaintiffs' attorney just mentioned, for whom Butcher

regularly testifies against the FBS. Butcher et al. (2008) discuss the content of these reviews in an effort to discredit the decision it yielded and offer to make all of the documents obtained by this attorney available upon request.

The reviews discussed by Butcher et al. (2008) were written by experts who had reason to believe that, as is customary, they were providing opinions strictly for the purpose of editorial review. That an attorney took advantage of the disclosure rules governing a public university should not give license to an expert working with her to violate this time-honored expectation. The unwarranted publication of excerpts from reviews written by experts with the reasonable expectation of privacy and with no intention that they be published and who did not authorize Butcher et al. (2008) to do so is an invasion of the editorial review process, which could have a chilling impact on the field. Faced with the prospect that their reviews will be published and made available to anyone upon request, how likely are reviewers to offer candid appraisals?

Because we do not wish to reinforce this conduct, we will not respond to the specifics of Butcher et al.'s (2008) analysis of the reviews. Suffice it to say that the review process followed standard procedures, resulted in a decision by the University of Minnesota Press to add the FBS to the standard scoring materials for the MMPI-2 based on the recommendations of a vast majority of the reviewers, and had withstood repeated challenges by Butcher and the attorney who provided him this material.

Butcher et al.'s Legal Analysis

Butcher et al. (2008) cite three Frye hearings in Florida's 13th Circuit and describe one judge's order excluding expert testimony based on FBS (*Williams v. CSX Transportation, Inc.* 2007). These authors are seemingly unaware that the overwhelming majority of courts in other jurisdictions allow evidence based on a variety of symptom validity techniques, even when the reliability and relevance of those techniques are directly challenged (e.g., *U.S. v. Bitton* 2008). Moreover, there are hundreds of cases in which expert testimony, based in part on FBS, is admitted without objection (e.g., *Mckinney-Prude v. Detroit Board of Education* 2007; *Moore v. Daimler Chrysler Corp.* 2007). The orders issued by Florida's 13th Circuit Court are isolated decisions demonstrating that judges are not always well informed about the extensive scientific evidence supporting FBS and other symptom validity techniques. Numerous board-certified clinical neuropsychologist experts report admission of FBS testimony into evidence with some testifying that they have never had FBS

excluded (e.g., *Upchurch v. Broward Co. School Bd.* 2008; *Solomon v. TK Power* 2008). In a recent FL case, objections to the FBS were withdrawn. Prior to the withdrawal, evidence and oral arguments that symptom validity techniques are reliable and generally accepted within the relevant scientific community were presented. In pending litigation, a FL judge hearing a Frye challenge to testimony based on FBS allowed the evidence to be admitted, but limited how it could be used to address the question of malingering (*Nason v. Shafranski* 2008).

In citing judicial decisions, Butcher et al. (2008) appear to be advancing a legal argument. However, lawyers are ethically obligated to acknowledge potentially adverse legal authority, as follows:

Rule 3.3(a)(2) Candor Toward The Tribunal. A lawyer shall not knowingly fail to disclose to the tribunal legal authority in the controlling jurisdiction known to the lawyer to be directly adverse to the position of the client and not disclosed by opposing counsel. (ABA Model Rules of Professional Conduct)

Although this rule does not apply to Butcher *et al.* in a journal article, their selective use of legal authority would not be accepted in a court of law. Indeed, if a lawyer made such unbalanced representations to a court, they would be subject to sanctions. Moreover, Butcher et al. (2008) offer no legal analysis of the decision and no objective description of the current legal standing of symptom validity techniques in our courts. As we have already stated, FBS is not recommended for use in isolation from other symptom validity techniques and observations to draw conclusions about malingering (Greiffenstein et al. 2007). The scale is just one of a growing set of techniques for detecting exaggerated symptoms, suboptimal effort, or noncredible performance during evaluations (hereinafter "SVT science"). SVT science is routinely admitted as evidence in various legal proceedings in an overwhelming number of state, federal, and international jurisdictions.

Notwithstanding Butcher et al.'s (2008) substantial overstatements and misuse of legal authority, the SVT science reviewed in this article is receiving more judicial scrutiny as it is being employed more routinely by experts. SVT science is on a collision course with evidence law and it requires courts to carefully review the rules governing experts and admissibility of expert testimony (Creager et al. 2002). The most common tactic used to restrict application of SVT science is a motion in limine in which an attorney asks the court to exclude SVT evidence from being heard by a jury. In civil cases, motions to exclude evidence are most commonly filed by plaintiff attorneys, while in criminal cases, the defendant is usually seeking to keep out evidence of malingering. Attorneys are advancing legitimate arguments for courts to consider regarding the

admissibility of SVT science. Although arguments to exclude SVT science may take a variety of forms, most are based on the rules of evidence and expert testimony, asserting that SVT science is: (1) more prejudicial than probative, (2) inadmissible character evidence, (3) wrongfully intruding into the province of the jury, or (4) not generally accepted by the relevant scientific community. We briefly address each to these arguments.

The first two arguments require an understanding of relevance as defined in Federal Rules of Evidence 401, as follows:

“Relevant evidence” means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

Essentially all relevant evidence is admissible, unless privileged. However, courts must balance other factors when determining the admissibility question as described in Federal Rule of Evidence 403, as follows:

Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

The judge in *Williams v. CSX Transportation, Inc.* (2007) weighed these concerns and determined, among other things, that the probative value of FBS was outweighed by its prejudicial effect, commenting that the term “faking bad” was overtly prejudicial³. In balancing the relevance of SVT science, this judge seemingly placed greater weight on the name of the scale rather than the reliability of its application.

The inadmissible character evidence argument is a derivative of the relevance question as addressed in Federal Rule of Evidence 404, as follows:

Evidence of a person’s character or a trait of character is not admissible for the purpose of proving action in conformity therewith on a particular occasion.

The rule has some complicated exceptions in criminal cases that are beyond the scope of this article, but in civil proceedings, character evidence is generally inadmissible unless character is at issue (e.g., defamation). When FBS is elevated at levels described in this paper, our best science indicates that the examinee was likely overendorsing

symptoms—a fact that plaintiff attorneys misconstrue as the expert calling the plaintiff a fake, a fraud, or a liar. As we will discuss below, Butcher’s testimonial support for this distortion is at odds with any reasonable application of the science on FBS.

In considering the best response to this inflammatory tactic, the testifying witness should remember that the scientific accuracy of their expert opinion and the confidence with which it is rendered may not necessarily translate into credibility with the trier of fact. The expert should always be mindful that jurisdictional restrictions, local customs, or judge idiosyncrasies may limit the scope of their opinions regarding symptom exaggeration or suboptimal effort. Experts should recognize that terms like fake, fraud, and liar when used in cross examination are drawing for character judgments in a transparent effort to impeach the credibility of the expert. So when the plaintiff attorney asks, “Are you calling my client a fake, fraud, and a liar?”, one effective response is to an answer in the negative and simply point out that an elevated FBS is just one indicator of symptom invalidity/intentional exaggeration associated with exaggerated symptoms and to allow the jury to draw its own conclusions after hearing all the evidence.

Respecting jurors conclusions is the basis for the third argument against FBS admissibility. Judges make decisions about admissibility of evidence, and generally, juries weigh the credibility of that evidence. In the end, the jury decides the credibility of the plaintiff’s claim, not an expert witness. Experts must express appropriate opinions within the scope of their expertise in a manner that is helpful to the jury (Federal Rule of Evidence 702). However, experts must not state legal conclusions that potentially invade the province of the jury. In this regard, Federal Rule of Evidence 704 is a source of confusion for some attorneys, as follows:

No expert witness testifying with respect to the mental state or condition of a defendant in a criminal case may state an opinion or inference as to whether the defendant did or did not have the mental state or condition constituting an element of the crime charged or of a defense thereto. Such ultimate issues are matters for the trier of fact alone.

Some attorneys misapply this rule in civil proceedings, while others overextend its reach by suggesting that experts cannot testify about their data when those data are directly relevant to a matter that a jury must decide, including whether or not symptoms or disabilities are exaggerated. In many respects, the “ultimate issue” rule is abandoned when the expert witness’ testimony is demonstrably helpful to the jury.

Having addressed the first three relevance-based arguments used in efforts to exclude FBS, the final argument questions the reliability of FBS. This strategy for excluding FBS uses the standards for evaluating experts as addressed

³ As noted earlier, in response to this concern, the name of the scale was subsequently changed to the more descriptive label “Symptom Validity”. The abbreviation FBS was retained to link the scale to its empirical literature.

in *Frye v. United States* (1923), *Daubert v. Merrill Dow Pharm. Inc.* (1993), and its progeny, Fed. R. Evid. 702 Testimony by Experts and Fed. R. Evid. 104 Preliminary Questions. Here, the judge plays the key role in determining reliability of the methods employed by expert witnesses. Briefly, a judge may deny the admission of evidence in a *Frye* jurisdiction, by simply finding that the methodology is not accepted in the relevant scientific community. The judge in *Vandergracht* (2005) made such a finding and excluded FBS, because there was not “ample evidence that the test is accepted by his peers.”

In their legal analysis, Butcher et al. (2008) quote from the judge’s decision in the most recent Florida case (*Williams v. CSX Transportation, Inc.* 2007) where FBS testimony was excluded:

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. (p. 11)

As is evident in his opinion, the judge in this case was presented with many of the same erroneous assertions that Butcher et al. (2008) advance in their current critique. The expert who provided this testimony is the lead author of the Butcher et al. (2008) critique. The following excerpts from this hearing illustrate the testimony upon which the Court relied in making its decision:

Q Okay. Let’s go to your criticisms. What concerns do you have about the Fake Bad Scale?

A The way in which it was constructed was not up to standards as far as test construction goes. And one of the major problems with the Fake Bad Scale is that it has a high false positive rate based upon the cutoff scores that were initially provided by Lees-Haley a cutoff score of 20.

And we published an article indicating that one of the main problems with the Fake Bad Scale—and this was conducted by Paul Arbisi and myself and a couple of other people—was that the Fake Bad Scale is comprised in large part of big chunks of items that are on existing symptom scales. So the same questions fall

on the Fake Bad Scale that are actually on mental health and health symptom scales. That’s the main problem with it.

Most of the research on the Fake Bad Scale has not really used malingerers, per se, but they’ve used litigants. And some litigants are not malingerers. Actually, many are not malingerers. And so that’s gotten kind of confused in the process.

The witness’s characterization of the scale as having a high false positive rate, in particular in reference to a cutoff that has long ago and repeatedly been identified by the developer of the scale as too low, is clearly at odds with the literature reviewed here. As we discussed earlier, the issue of item overlap with substantive scales is a red herring; the same criticism could be leveled at this witness’s favored scale, F, and would be similarly misleading. The testimony that most of the research on the FBS has not used real malingerers is factually incorrect and inconsistent with the literature reviewed earlier (see for example Bianchini et al. 2008 and Greve et al. 2006a, b).

Q It says, “Score of 22 or higher.” So, for example, if somebody gets a 23 and they’re a woman, what percent of those individuals in your sample have you found to be malingering?

A If you look at just 22 and higher—

Q Uh-huh.

A —44 percent of women would be considered malingering in an inpatient psychiatric setting; they’re in there for treatment, and they would be considered malingering.

Here, the witness demonstrates Butcher et al.’s (2008) erroneous equating of elevated scores on FBS with malingering and compounds the misleading nature of the testimony by relying on a cutoff lower than the one recommended by Greiffenstein et al. (2007) for interpreting scores of women with a history of psychiatric disorder. Moreover, the data are those reported by Butcher et al. (2003) where no information was available on whether these test-takers had any incentive to overreport a necessary condition for a finding of malingering.

In response to a question about modifying cutoffs for FBS interpretation, the witness stated:

A He has—he has altered his cutoff standard based on a number of things, including their most recent study of the 2007 article, they’ve jumped it way up to—and it’s a—it’s a variable standard.

For example, I think, they call for a 26 cutoff is recommended for someone that has chronic and severe

brain damage, or they recommend 29 plus if there's some kind of pre. For women if there's a some kind of a pre-injury psyche history, or 30 plus is recommended for those with a medical history that's complex and so forth.

So, there is not a single cut score in the literature. It's wherever you look, you see a different picture.

Here, the witness demonstrates that he is indeed aware that the cutoff he referred to in the previous excerpt is incorrect and inappropriate. Moreover, as we indicated earlier, contrary to the impression generated by this testimony, modifying cutoffs for MMPI-2 validity scales is a standard practice. For example, the MMPI-2 manual (Butcher et al. 2001) recommends different cutoffs for identifying overreporting based on the F scale for nonclinical, clinical outpatient, and clinical inpatient settings.

As discussed earlier, Butcher et al. (2008) insinuate a connection between feedback provided by the developer of the FBS on a preliminary set of scales for a new version of the MMPI-2 and the addition of the scale to the MMPI-2 scoring materials. A similar attempt by the witness and the plaintiff's attorney in the *Williams* Frye hearing was rebuffed by the judge:

Q Can you tell me, sir, whether or not you're aware of the University of Minnesota Press through Dr. Ben-Porath deciding to include the MMPI scale—Fake Bad Scale created by Dr. Paul Lees-Haley, and also a letter from Dr. Paul Lees-Haley just before that acceptance recommending the use of Dr. Ben-Porath's shorter test forms?

Mr. F: Objection, Your Honor, leading.
THE COURT: Overruled.

Q Go ahead.

A That's correct.

Q Doctor, do you have an opinion as to whether or not there is any quid pro quo or potential for quid pro quo involved in something like that, you approve my scales and I'll approve yours?

MR. F: Objection, Your Honor.
MS. S: Let me ask it another way.

Q Can you rule it out?

MR. F: Objection, Your Honor, compound, leading.
THE COURT: Sustained.

Finally, the witness offered the judge in the *Williams* case this observation:

A In my view, they present the Fake Bad Scale as like a silver bullet that goes into the person's psyche and picks out malingering, when in my personal view, in my opinion, it's more like a crude improvised

explosive device that blows everything up. And that's the way these folks are using the test. When they see that FBS up, the person is malingering, there's nothing else to say.

Such inflammatory language reveals a personal bias that serves neither the scientific community in its efforts to assess the validity and utility of FBS nor the legal community's need to rely on objective experts in understanding the scientific literature. Along the same lines, in an interview this witness gave to the *Wall Street Journal* he stated in reference to FBS "virtually everyone is a malingerer according to this scale. This is great for insurance companies but not great for people" (Armstrong 2008, March 5).

As these excerpts reflect, the judge's opinion in the *Williams* case was swayed by testimony that is inconsistent with the scientific literature and characterized by many of the same flaws we have demonstrated here in the Butcher et al. (2008) article. Rather than providing confirmation of the accuracy of Butcher et al.'s (2008) critique, the *Williams* decision reflects the problems trial judges face when presented with misleading testimony.

Frye Versus Daubert

Although not applicable in the isolated *Frye* rulings that excluded FBS in a few Florida cases, the *Daubert* analysis is more complex and is applied in all federal courts and a majority of states. *Daubert* examines whether the theory and methods used (1) were generally adopted by the scientific community (*Frye* "general acceptance" test), (2) were subject to peer review and publication, (3) could be or had been tested, and (4) has a known and acceptable error rate (*Daubert*, p. 597). Although these factors are not exclusive, most courts apply them to determine the admissibility of evidence. There is not a single case of FBS failing a *Daubert* challenge.

In 2002, holdings from the *Daubert* and its progeny were used to amend Rule 702 and codify these US Supreme Court decisions into the current rules governing expert testimony. Rule 702 reads as follows:

"If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case." (FED. R. EVID. 702)

Although FBS challenges will continue, when SVT science is presented to most courts, FBS testimony will be found to be based on sufficient data that is the product of reliable methods by experts who appropriately apply those methods reliably to the facts of a case.

Courts will scrutinize FBS and other symptom validity techniques, probing the relevance and reliability of each methodology. However, Butcher et al. (2008) have not made a persuasive scientific or legal case against FBS. SVT science will survive its collision with evidence law and the neuropsychologist expert using FBS along with other techniques will assist the jury in resolving questions of credibility of claims.

Summary

We scrutinized Butcher et al.'s (2008) criticisms of the FBS and identified major conceptual, methodological, and empirical flaws in their arguments. These authors incorrectly equate a positive finding on the FBS with a diagnosis of malingering, which runs counter to current recommended practices in the field in general and interpretive recommendations for the scale in particular. Absent evidence of an external incentive and other (than the FBS score) indications of overreporting, none of the individuals that Butcher et al. (2008) claim would have been “diagnosed as malingering”, would actually have been so classified. In addition, these authors apply cutoffs that are lower than the ones currently recommended for the scale to samples of individuals who were not screened for extratest evidence of overreporting yielding an indeterminable proportion of subjects who could have been overreporting and thus rendering their reported “false positive” rates uninterpretable.

We also examined Butcher et al.'s (2008) criticisms of FBS in light of their views, recommendations, and practices regarding other MMPI-2 scales. We found that their examples and illustrations include an erroneous conceptual foundation of the F scale, ignore the nature of the “norms” for this scale and years of interpretive practices with the original L and F scales of the MMPI, provide a misleading analysis of item overlap between FBS and substantive MMPI-2 scales that fails to consider the implications of similar overlap between their favored scale, F, and substantive measures of thought dysfunction, and reflect a double standard regarding “high stakes” psychological measures and assessments.

Next, we examined Butcher et al.'s (2008) review of the literature on FBS and showed that it fails to adequately reflect the considerable evidence of the scale's validity as a measure of overreporting in personal injury litigants and claimants. We specifically addressed their claims that the

scale is “biased” against individuals with disabilities and women and showed that both the existing literature and previously unreported data demonstrate that when properly interpreted, scores on FBS do not differentially identify individuals with disabilities or women as possibly overreporting let alone malingering.

Finally, we considered Butcher et al.'s (2008) analysis of a recent legal decision to exclude FBS testimony in a case heard in Florida. We identified significant problems with their legal analysis and showed that the judge in this case ruled on the basis of the same misleading information contained in their critique; therefore, rather than supporting Butcher et al.'s (2008) views, the court's decision in this case reflects the negative impact that misleading testimony can have on the judicial process.

In closing, we have shown that there is a solid empirical foundation for the clinical and forensic use of the FBS. Despite Butcher et al.'s (2003, 2008) arguments, there is no true scientific controversy. Indeed, the views and arguments presented by Butcher et al. (2008) are inconsistent with current research findings and practice recommendations as well as the conclusions of authorities in the field who have no direct involvement with this research. For example, after reviewing the literature on FBS (including Butcher et al.'s 2003 arguments) for the third edition of their *Compendium*, Strauss et al. (2006) concluded:

Although the value of the FBS to detect suboptimal effort has been questioned (Butcher et al. 2003), the available evidence suggests that it provides unique information over and above traditional MMPI-2 validity indices in personal injury cases, including exaggerated pain, post-traumatic anxiety, and neurological problems. (p. 1123)

Butcher et al.'s (2003) criticisms stimulated some of the subsequent research aimed at better characterizing the validity of FBS. However, that research has repeatedly failed to support their conclusions and recommendation against using the scale. Rather, thoughtfully designed and well-conducted studies have consistently demonstrated the valuable role of FBS in forensic psychological and neuropsychological assessment.

Disclosure Yossef Ben-Porath is a paid consultant to the publisher of the MMPI instruments, the University of Minnesota Press, and their distributor, Pearson. He receives royalties on sales of the MMPI-2-RF.

References

- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, *21*, 899–916.

- Armstrong, D. (2008). Malingering test roils personal-injury law: "Fake Bad Scale" bars real victims, its critics contend. *The Wall Street Journal*, p. A1.
- Barr, W. B. (2005). *Rates of invalid MMPI-2 responding in patients with epileptic and nonepileptic seizures*. Poster session presented at the American Epilepsy Society annual meeting, Washington, DC.
- Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society*, *11*, 215–27.
- Ben-Porath, Y.S. and Tellegen, A. (2007). MMPI-2 FBS (Symptom Validity) Scale. Retrieved June 16, 2008 from <http://www.pearsonassessments.com/resources/fbs.htm>.
- Ben-Porath, Y. S., Tellegen, A., & Graham, J. R. (2009). *The MMPI-2 Symptom Validity Scale (FBS)*. Minneapolis, MN: University of Minnesota Press.
- Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: A dose–response relationship? *The Clinical Neuropsychologist*, *20*, 831–847.
- Bianchini, K. J., Etherton, J. L., Greve, K. W., Heinly, M. T., & Meyers, J. E. (2008). Classification accuracy of MMPI-2 Validity Scales in the detection of pain-related malingering: A known-groups approach. *Assessment*, *15*, 435.
- Bianchini, K. J., Greve, K. W., & Love, J. M. (2003). Definite malingered neurocognitive dysfunction in moderate/severe traumatic brain injury. *The Clinical Neuropsychologist*, *17*, 574–580.
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, *5*, 404–417.
- 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.
- Binder, L. M., Rohling, M. L., & Larrabee, G. (1997). A review of mild head trauma. Part I: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology*, *19*, 421–431.
- Bury, A. S., & Bagby, R. M. (2002). The detection of feigned uncoached and coached posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment*, *14*, 472–484.
- Butcher, J. N. (1998). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2) user's guide for the Minnesota Report: Reports for forensic settings*. Bloomington, MN: Pearson Assessments.
- Butcher, J. N. (2001). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2) user's guide for the The Minnesota Report: Revised personnel system* (3rd ed.). Bloomington, MN: Pearson Assessments.
- 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, NJ: Erlbaum.
- Butcher, J. N., & Williams, C. L. (2000). *Essentials of MMPI-2 and MMPI-A interpretation* (2nd ed.). Minneapolis, MN: University of Minnesota Press.
- Butcher, J. N., & Perry, J. N. (2008). *Personality assessment in treatment planning: Use of the MMPI-2 and BTPI*. New York, NY: Oxford University Press.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota multiphasic personality inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2): Manual for administration, scoring, and interpretation, revised edition*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Arbisi, P. A., Atlas, M. M., & McNulty, J. L. (2003). The construct validity of the Lees-Haley Fake-Bad Scale (FBS): Does this scale measure somatic malingering and feigned emotional distress? *Archives of Clinical Neuropsychology*, *18*, 473–485.
- Butcher, J. N., Gass, C. S., Cumella, E., Kally, Z., & Williams, C. L. (2008). Potential bias in MMPI-2 assessments using the Fake Bad Scale (FBS). *Psychological Injury and the Law*, *1*, 191–209.
- Carroll, L. J., Cassidy, J. D., Peloso, P. M., et al. (2004). Prognosis for mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine*, (43 Suppl), 84–105.
- Chafetz, M. D. (2008). Malingering on the social security disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist*, *22*, 529–546.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Creager, R. T., Shea, J. C., & Larner, G. P. (2002). Role of defense neuropsychologists should be limited under Virginia evidence law. *The Journal of the Virginia Trial Lawyers Association*, *14*, 24–32, Fall.
- Daubert v. Merrill Dow Pharm., Inc.*, 509 U.S. 589–95 (U.S. 1993).
- Frye v. United States*, 509 U.S. 579 (D.C. Cir. 1923).
- Gallop, R. J., Crits-Christop, P., Muenz, L. R., & Tu, X. M. (2003). Determination and interpretation of the optimal operating point for ROC curves derived through generalized linear models. *Understanding Statistics*, *2*, 219–242.
- Geisinger, K. F. (2005). The testing industry, ethnic minorities, and individuals with disabilities. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 187–203). Mahwah, NJ: Erlbaum.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Greiffenstein, M. F., & Baker, W. J. (2001). Comparison of premorbid and postinjury MMPI-2 profiles in late postconcussion claimants. *Clinical Neuropsychologist*, *15*, 162–170.
- Greiffenstein, M. F., Baker, W. J., Axelrod, B., Peck, E. 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., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 fake bad scale in detection of noncredible brain injury claims. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 210–235). New York, NY: Guilford.
- Greve, K. W., & Bianchini, K. J. (2004a). Setting empirical cut-offs on psychometric indicators of negative response bias: a methodological commentary with recommendations. *Archives of Clinical Neuropsychology*, *19*, 533–541.
- Greve, K. W., & Bianchini, K. J. (2004b). Response to Butcher et al., The construct validity of the Lees-Haley Fake-bad scale. *Archives of Clinical Neuropsychology*, *19*, 337–339.
- Greve, K. W., Bianchini, K. J., & Ameduri, C. J. (2003). The use of a forced-choice test of tactile discrimination in the evaluation of functional sensory loss: A report of 3 cases. *Archives of Physical Medicine and Rehabilitation*, *84*, 1233–1236.
- Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., et al. (2006a). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *NeuroToxicology*, *27*, 940–950.
- Greve, K. W., Bianchini, K. J., Love, J. M., Brennan, A., & Heinly, M. T. (2006b). Sensitivity and specificity of MMPI-2 Validity Scales and indicators to malingered neurocognitive dysfunction in traumatic brain injury. *The Clinical Neuropsychologist*, *20*, 491–512.
- Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and

- chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist*, 22, 896–918.
- Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (in press). The prevalence of malingering in chronic pain patients referred for psychological evaluation in a medico-legal context. *Archives of Physical Medicine & Rehabilitation*.
- Guez, M., Brannstrom, 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.
- Harris, I., Mulford, J., Solomon, M., van Gelder, J. M., & Young, J. (2005). Association between compensation status and outcome after surgery: A meta-analysis. *Journal of the American Medical Association*, 293, 1644–1652.
- Hathaway, S. R., & McKinley, J. C. (1942). A multiphasic personality schedule (Minnesota): III. The measurement of symptomatic depression. *Journal of Psychology*, 14, 73–84.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota multiphasic personality inventory* (Rev. ed., 2nd printing). Minneapolis, MN: University of Minnesota Press.
- Hathaway, S. R., & McKinley, J. C. (1951). *Minnesota multiphasic personality inventory; manual* (Revised). San Antonio, TX: Psychological Corporation.
- Iverson, G. L. (2003). Detecting malingering in civil forensic evaluations. In A. M. Horton, & L. C. Hartlage (Eds.), *Handbook of forensic neuropsychology* (pp. 137–177). New York: Springer.
- Iverson, G. L., & McCracken, L. M. (1997). 'Postconcussive' symptoms in persons with chronic pain. *Brain Injury*, 11, 783–790.
- Kay, N. R., & Morris-Jones, H. (1998). Pain clinic management of medico-legal litigants. *Injury*, 29, 305–308.
- 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. (2003). Exaggerated MMPI-2 symptom report in personal injury litigants with Malingered Neurocognitive Deficit. *Archives of Clinical Neuropsychology*, 18, 673–686.
- Larrabee, G. J. (2005). Assessment of malingering. In G. J. Larrabee (Ed.), *Forensic Neuropsychology* (pp. 115–158). 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. 334–371). 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.
- Mckinney-Prude v. Detroit Board of Education*, (May 16, 2007 Order) Workers Compensation Board of Magistrates.
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, 30, 525–564.
- Meyers, J. E., Millis, S. R., & Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology*, 17, 157–169.
- Meyers, J. E., & Rohling, M. L. (2004). Validation of the Meyers Short Battery on mild TBI patients. *Archives of Clinical Neuropsychology*, 19, 637–651.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102.
- Moore v. Daimler Chrysler Corp.*, (February 16, 2007 Order) Workers Compensation Board of Magistrates.
- Nason v. Shafranski*, (September 11, 2008 Ruling), Florida 19th Circuit, St. Lucie County.
- Nelson, N. W., Parsons, T. D., Grote, C. L., Smith, C. A., & Sisung 2nd, J. R. (2006). The MMPI-2 Fake Bad Scale: concordance and specificity of true and estimated scores. *Journal of Clinical and Experimental Neuropsychology*, 28, 1–12.
- Pope, K. S., Butcher, J. N., & Seelen, J. (2006). *MMOI/MMPI-2/MMPI-A in court: Assessment, testimony, and cross-examination for experts and attorneys*. Washington, DC: American Psychological Association.
- Public Attitude Monitor. (1992). (Survey). Oak Brook, IL: Insurance Research Council.
- Public Attitude Monitor. (1993). (Survey). Oak Brook, IL: Insurance Research Council.
- Rainville, J., Sobel, J. B., Hartigan, C., & Wright, A. (1997). The effect of compensation involvement on the reporting of pain and disability by patients referred for rehabilitation of chronic low back pain. *Spine*, 22, 2016–2024.
- Rohling, M. L., Binder, L. M., & Langhinrichsen-Rohling, J. (1995). A meta-analytic review of the association between financial compensation and the experience and treatment of chronic pain. *Health Psychology*, 14, 537–547.
- Ross, S. R., Millis, S. R., Krukowski, R. A., Putnam, S. H., & Adams, K. M. (2004). Detecting incomplete effort on the MMPI-2: an examination of the Fake-Bad Scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology*, 26, 115–124.
- 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.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingering neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561.
- Solomon v. TK Power*, Case 06-CA-003088, Florida 4th Circuit, Duval County.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). NY: Oxford University Press.
- Su, J. C., & Birmingham, C. L. (2003). Anorexia nervosa: The cost of long-term disability. *Eating and Weight Disorders*, 8, 76–79.
- Tsanadis, J., Montoya, E., Hanks, R. A., Millis, S. R., Fichtenberg, N. L., & Axelrod, B. N. (2008). Brain injury severity, litigation status, and self-report of postconcussive symptoms. *The Clinical Neuropsychologist*, 22, 1080–1092.
- 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.
- Upchurch v. Broward Co. School Bd.* (2008) January 24, 2008 Deposition of Dr. James N. Butcher, Florida 15th Circuit, Broward County.
- United States v. Bitton* (July 1, 2008 Memorandum Decision and Order) Case No. 2:05-CR-661 TS United States District Court of Utah, Central Division
- Vaccaro, A. R., Ring, D., Scuderi, G., Cohen, D. S., & Garfin, S. R. (1997). Predictors of outcome in patients with chronic back pain and low-grade spondylolisthesis. *Spine*, 22, 2030–2034.
- Vandergracht v. Progressive Express* No. 02-04552 (Fla. Cir. Ct. Hillsborough County, March 9, 2005).
- Williams v. CSX Transportation, Inc.* (Sept 19, 2007 Order).