

Inaccuracies About the MMPI-2 Fake Bad Scale in the Reply by Ben-Porath, Greve, Bianchini, and Kaufman (2009)

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

Received: 11 January 2009 / Accepted: 12 April 2009 / Published online: 15 May 2009
© Springer Science + Business Media, LLC 2009

Abstract Based on a focused review and careful analysis of a large amount of published research, Butcher et al. (Psychol Inj and Law 1(3):191–209, 2008) concluded that the Fake Bad Scale (FBS) does not appear to be a sufficiently reliable or valid measure of the construct

“faking bad”. Butcher et al. (Psychol Inj and Law 1(3): 191–209, 2008) pointed out examples of errors in some of the most widely cited studies (including meta-analytic) used to support the FBS and described potential biases if the FBS is used to impute the motivation to malingering in

Disclosure statement: Butcher, Williams, and Cumella worked on the development of the MMPI-2 and/or MMPI-A; none receives royalties from sales of those instruments or their scales. None of the authors received consultant fees or other personal income for their contributions to this article. Butcher is the author of an interpretive system for the MMPI-2, the Minnesota Report. The author royalties for the Minnesota Reports are 30% of the royalties the University of Minnesota receives from Pearson Assessments from its net sales and are well in excess of \$10,000 annually. Butcher receives the full amount for the personnel and forensic systems, shares half with his coauthor of the adolescent system (Williams), and voluntarily assigned 5% to Williams for her consultation on the Adult Clinical System. His part time practice includes forensic cases and he has testified fairly evenly for plaintiffs and defense. Since 1996, he has had nine cases involving the Fake Bad Scale, four in 2008 (two pro bono). Williams splits the author royalties from the Minnesota Reports with Butcher as described above. She is a consultant to Butcher’s practice and receives income from that. Her work does not include expert witness testimony. Disclosure statements for Butcher and Williams appear at <http://www1.umn.edu/mmpi/disclosure.php>. Gass derives 10% to 15% of his total professional income from his private practice and approximately 75% of this activity is based on defense work. Cumella and Kally are full-time employees at Remuda Program for Eating Disorders and do not receive other compensation related to the content of this article.

C. L. Williams (✉) · J. N. Butcher
University of Minnesota,
Minneapolis, MN, USA
e-mail: willi001@umn.edu

J. N. Butcher
e-mail: butch001@umn.edu

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

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

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

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

those reaching its variable and imprecise cutoff scores. In a response to this article, Ben-Porath et al. (Psychol Inj Law 2:62–85, 2009) dismissed all the concerns raised in it with suggestions that our conclusions were based on faulty premises, misunderstandings of basic concepts, misleading descriptions of Minnesota Multiphasic Personality Inventory (MMPI) research, flawed analyses, and so on. This reply corrects some of Ben-Porath and colleagues' (Psychol Inj Law 2:62–85, 2009) multiple misrepresentations of the points made in Butcher et al. (Psychol Inj and Law 1(3):191–209, 2008) and identifies eight logical fallacies relevant to the FBS controversy. We end with a challenge to other psychologists to fully examine the underlying FBS research before adopting this scale in their clinical practice.

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

Ben-Porath et al. (2009) response to our delineation of some of the potential biases of using the Fake Bad Scale (FBS) is largely based on a misrepresentation of, or failure to respond to, key aspects of our critique. Our article was repeatedly criticized for use of logical fallacies such as *straw person* or *red herring* arguments. We disagree with the inaccurate characterizations of our work in their response and welcome this opportunity to provide further perspective to other professionals in order that they can come to their own informed conclusions about using the FBS to raise questions about the veracity of individuals' self-report on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and/or performance on measures of neurocognitive functioning.

As our original article (Butcher et al. 2008) and the response (Ben-Porath et al. 2009) illustrate, use of the FBS is highly controversial among MMPI-2 experts, and given this, practitioners should fully evaluate this measure before using it with their clients, rather than simply accepting one side's opinion. Unfortunately, as we pointed out in our critique, this controversial scale was included in the widely used MMPI-2 Extended Score Report with only limited guidance posted on a website about how to avoid its misuse (i.e., Ben-Porath and Tellegen 2007a, b; Pearson Assessments 2007a).

Two years later, there is still no test manual or manual supplement with information to assist in psychologists' understanding of the use and limitations of the FBS. Even basic information such as FBS item membership and scoring directions, T score conversion tables, and endorsement frequencies by gender is unavailable in one basic source. While Ben-Porath et al. (2009, p. 65) promise that a "forthcoming MMPI-2 test monograph will provide T score

conversion tables and interpretive recommendations for FBS expressed in T scores," it should go without saying that a forthcoming publication is of no utility to psychologists or their patients until it is available.

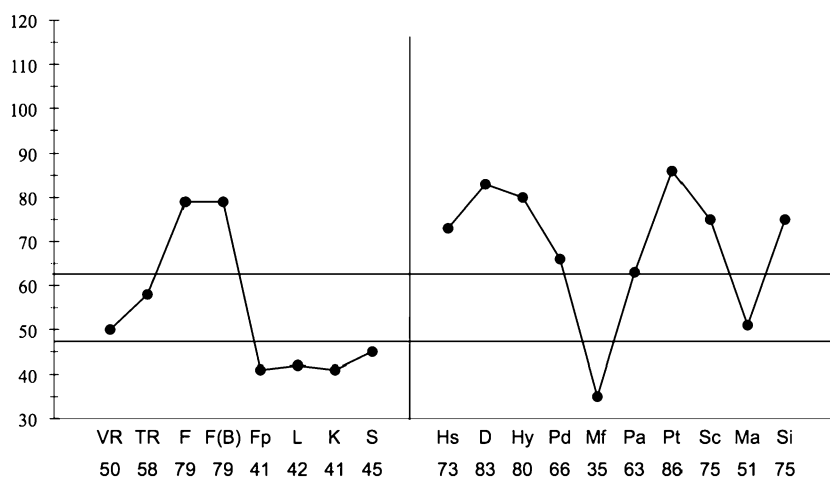
We begin our response with a case illustration to highlight the potential consequences of psychologists using the FBS in their clinical evaluations with the current publisher's guidelines (Ben-Porath and Tellegen 2007a, b) and without careful consideration about the potential biases we described in Butcher et al. (2008). Next, we respond to the criticism that we misunderstood the research on malingering and how to use the FBS. We provide more background for the name change of the FBS than what was included in Ben-Porath and colleagues' (2009) response. We reiterate our concerns about the inherent gender bias in the FBS and respond to Ben-Porath and colleagues' (2009) comments about our criticisms of the review process that led to the inclusion of the FBS in the MMPI-2. We respond to their claims that our article included logical fallacies and identified eight logical fallacies (i.e., *appeal to authority*, *appeal to emotion*, *straw person*, *red herring*, *ad hominem*, *false analogy*, *cherry picking*, and the *psychologist's fallacy*) that are related to the FBS and arguments included in their response (see Pope 2009 for his list of logical fallacies that occur frequently in psychological research—half of the eight we identified are included in Pope's 21 types). We conclude with a response to their *ad hominem* assertions and misrepresentations about our four-paragraph summary of the three *Frye* challenges in Florida courts concerning FBS use in expert witness testimony and include descriptions of three other *Frye* hearings, two in Florida and one in California, as well as a jury trial in California. We note, however, because of the short deadline given to us by the journal for this reply and a desire to be succinct that this is not an exhaustive detailing of all of our concerns about the response by Ben-Porath et al. (2009) to our article on the potential for biases with the FBS (Butcher et al. 2008).

A Case Illustration of the Potential for Harm from Use of the FBS

A licensed psychologist with 10 years postdoctoral experience administered the MMPI-2 as a routine part of an intake assessment of an inpatient with bulimia nervosa, obtaining the Extended Score Report from Pearson Assessments. The psychologist noted that the patient scored 29 on the FBS. Figure 1 contains the MMPI-2 validity and clinical scales for this patient.¹ Although all other MMPI-

¹ Experts on MMPI-2 profiles from patients with eating disorders will recognize this profile as fairly typical. See Cumella et al. (2000).

Fig. 1 MMPI-2 Validity and Clinical Scales Profile of an eating disorder in patient with FBS raw score of 29



2 validity scales indicated a valid profile for this patient, the psychologist, wishing to provide an accurate and comprehensive report of the patient's symptoms and issues, explained to the patient's treatment team that the patient should be assessed further for possible malingering. In the psychological evaluation, the psychologist quoted the following publisher guideline as justification for suggesting possible malingering:

“... the experts supporting addition of the FBS to the standard set of MMPI-2 validity scales recommended that raw scores above 22 should raise concerns about the validity of self-reported symptoms and that raw scores above 28 should raise very significant concerns about the validity of self-reported symptoms, particularly with individuals for whom relevant physical injury or medical problems have been ruled out” (Ben-Porath and Tellegen 2007a).

As a result of this psychologist's evaluation, the following events ensued that were complex and harmful to the patient. First, the treatment team, doubting its initial assessment of the patient's needs, entered a phase of reassessment, during which the patient was effectively deprived of needed treatment interventions for 2 days. Next, the patient's healthcare insurance company conducted a routine review of inpatient necessity during which its representative read the psychologist's evaluation in the patient's record. Based on the results of the FBS reported in the evaluation, the company denied further inpatient benefits for the patient. The patient's spouse, having witnessed firsthand the patient's severe bulimic symptoms and knowing the patient's need for inpatient care, agreed to pay for the expensive treatment out of pocket so that the patient could remain hospitalized. Nevertheless, the course of treatment was further impeded by the patient's knowledge of the financial stress that this decision imposed on her family.

Fortunately, the treatment team recognized the need for expert consultation in this case and informed management level clinicians at the inpatient agency of the issues that had arisen with the patient. A management level psychologist reviewed the case, was familiar with the controversies in the professional literature about the FBS, especially the evidence about the ambiguity of FBS cutoff scores, and determined that the FBS result, standing alone, was insufficient to suggest that this patient was malingering. The treatment team thus righted itself and resumed appropriate interventions with the patient. The assessing psychologist dictated an addendum to the original psychological evaluation and contacted the patient's insurance company to inform its representatives of the correction. Again, fortunately, the company's initial decision to deny insurance benefits was reversed, so that patient and family were not unduly burdened with the costs of inpatient care that rightly qualified for coverage within the benefits of their insurance policy. Agency management extended the patient's length of stay by 3 days, free of charge, to allow the patient additional time to engage in the needed treatment for severe bulimic symptoms and to compensate for the compromised treatment days. As such, in the long run, the initial harm that had occurred to the patient was remediated.

To be sure, the assessing psychologist did not undertake a careful review of the literature on the FBS before offering a possible interpretation to the patient's treatment team. What this well-intentioned psychologist did do, however, was to rely in good faith on the tools provided by the test distributor's Extended Score Report to facilitate interpretation of the MMPI-2. The psychologist trusted the judgment of the publisher, distributor, and its experts that a patient's score that exceeded the recommended cutoff of 28 raised “very significant concerns about the validity of self-reported symptoms” and included this information in the

patient's psychological evaluation. Psychologists currently do not have either a test manual or supplement with comprehensive information about this scale, including information about the controversies in the field regarding its use.

It is likely that this psychologist is not unique, but represents other psychologists who operate in good faith and rely on the tools of their profession, especially well-respected tools such as the MMPI-2 and its test distributor's Extended Score Report. As such, this case is illustrative of the type of harm that may occur to other patients as a result of FBS inclusion in standard MMPI-2 reports. Our concern is that not every case is likely to resolve as satisfactorily as the one herein described. Psychologists, especially given their past experiences with the MMPI-2 as a well-validated measure, may be vulnerable to the *appeal to authority* fallacy (i.e., an assertion is deemed true because the authorities say it is so) if they adopt the FBS into their clinical practices without a careful understanding of its underlying research support.

Understanding Malingering and the Use of the FBS

According to Ben-Porath et al. (2009), 40% of Americans believe that it is acceptable to make a "purposeful misrepresentation" (i.e., lie) about compensation claims and "studies suggest that the rate of malingering likely ranges from 20% to 50% across a wide range of clinical conditions and medicolegal contexts" (p. 64). Clearly, if their assertions are correct and if the MMPI-2 item pool could be used to develop a scale to reliably distinguish between malingerers and those who fall into other diagnostic categories, such a scale would have platinum status. However, for the multiple reasons and empirical studies we reviewed in our article, we do not believe that the FBS is a reliable tool for identifying malingering and question whether the MMPI-2 item pool has the necessary content to measure this construct.

One of their criticisms of our article is that we misunderstood research on the diagnosis of malingering and the specific recommendations for interpretations of FBS scores. Ben-Porath et al. (2009) assert that the FBS was not designed as a single index of malingering. This assertion, however, is not correct; the FBS was developed exclusively for this purpose, as clearly indicated by Lees-Haley et al. (1991, p. 203): "This paper presents a scale for using the Minnesota Multiphasic Personality Inventory-2 for the detection of malingers in personal injury claims." Lees-Haley et al. (1991, p. 204) continued:

"Malingering is a serious problem in the evaluation of patients who are involved in making claims for

financial compensation or disability leave, especially those claims in litigation. Representatives of litigating patients deliberately exaggerate damages rather than attempting to present scientifically accurate assessments of plaintiff's damages. For example, attorneys openly admit that "they ask for more than we expect to get" in making their demands, perhaps in part based on the belief that their adversaries expect a negotiating process to ensue. Attorneys representing plaintiffs are taught courses in how to "maximize damages." They deliberately suppress clinically important data which might interfere with their goals."

As we pointed out in our article, subsequent studies of the FBS demonstrated the unexpected findings of a closer relationship of FBS scores with clinical scales measuring somatic symptoms and somatoform disorder as opposed to the established MMPI-2 measures of symptom exaggeration (i.e., F, F_b, F_p; see Butcher et al. 2003). Proponents argue that the FBS assesses a different type of malingering than the F family, especially useful in forensic situations such as personal injuries (Greiffenstein et al. 2007; Larrabee 1998, 2007). The widely cited study by Larrabee (1998) is especially illustrative with his conclusion that the FBS is a better measure of "somatic malingering" because it identified 11 of 12 individuals in his convenience sample as malingering, whereas the F Scale "only" identified three of his convenience sample as malingering. So, despite the contrary assertions by Ben-Porath et al. (2009), the FBS was developed as a single measure of malingering and is intended to be interpreted as indicative of symptom exaggeration even in the absence of elevations on other MMPI-2 validity scales and even when the patient produces a clinical profile consistent with his or her presenting complaints, as occurred in our case illustration (Fig. 1).

Furthermore, while Ben-Porath et al. (2009, p. 63) indicate that a "positive FBS score alone is insufficient for diagnosis," specifying that the presence of some external incentive is necessary that advice does not appear in the web-based guidelines (i.e., Ben-Porath and Tellegen 2007a, b) posted for clinicians to rely on in their use of the FBS. That this advice appears in a chapter recommended to test users (e.g., Greiffenstein et al. 2007) does not provide sufficient guidance for psychologists using the scale, particularly when the chapter was in press at the time the FBS was added to the Extended Score Report. This again highlights the need for a comprehensive resource on the FBS for practitioners, one that accurately conveys its limitations.

In an apparent contradiction of their recognition that an external incentive for malingering is necessary before the FBS should be utilized, Ben-Porath and Tellegen (2007a, b) recommended the FBS for all settings in which the MMPI-

2 is used, even those with low probability of secondary gain from symptom reports, as in the eating disorders sample described in our original article. Secondary gain for patients has been defined as “an external incentive to prolong symptom reporting beyond reasonable recovery times”. Examples include compensation, attendant care services for family members unable to find better paying work, access to narcotic medications, and societal forgiveness of the adult expectations to work (Lamberty 2008, p. 51).

As we pointed out in our article, Lees-Haley et al. (1991) hypothesized that a substantial number of clinicians, in addition to attorneys, coach claimants in advance of independent medical examinations to create false claims. It should be pointed out that in a personal injury setting, patients and their psychologists and attorneys are not the only parties with the potential for secondary gain. Defense attorneys for industry and their insurance carriers, as well as their psychological consultants and expert witnesses, are also subject to the possibility of conflicts of interest. Young (2008) in describing the conflict of interest policy of this journal, *Psychological Injury and Law*, noted that the type of work typically undertaken in a psychologist’s practice (i.e., plaintiff or patient-related, defense or third-party payer-related, or both) could be a source of potential conflicts of interest that must be disclosed by authors submitting articles to this journal. Other areas for potential conflicts of interest requiring disclosure include “all professional links that authors of submissions may have with advocacy representatives, liaisons, pharmaceutical companies, test companies, or other commercial ventures or organizations related in any way to the submission in question . . .” (Young 2008, pp. 5–6). Full and complete disclosure is one method for managing potential conflicts. The authors of this article have included such disclosures in a footnote.²

The Name Change

Ben-Porath et al. (2009) are now referring to the FBS as the Symptom Validity Scale (FBS). This name change speaks volumes and is a very rare occurrence in the history of psychological measurement. It is tantamount to an admission that the FBS is not a specific measure of conscious faking (i.e., fake bad) and that the scale authors (Lees-Haley et al. 1991) confounded malingering with other sources of symptom reporting during its initial development. Furthermore, since 1991, all publications about this

measure used the name Fake Bad Scale, and it was added to the MMPI-2 Extended Score Report with that name (e.g., Ben-Porath and Tellegen 2007a, b; Pearson Assessments 2007a). In a footnote, Ben-Porath et al. (2009, p. 62) stated:

“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.”

The FBS was added to the MMPI-2 in January 2007. Almost a year later on September 3 when we downloaded the January press release from the test publisher (Pearson Assessments 2007a) and accompanying statements from Ben-Porath and Tellegen (2007a, b), its name was still “Fake Bad Scale”. On September 19, 2007, Judge Bergmann, ruling in *Williams v CSX Transportation* (2007, p. 12), indicated:

“The very name ‘Fake Bad Scale’ is pejorative and derogatory and thus prejudicial.”

Following this ruling sometime in November or December 2007, the Pearson Assessments web-based statements about the FBS had the scale’s new name without any explanation or indication that the web statements had been altered (e.g., Ben-Porath and Tellegen 2007c; Pearson Assessments 2007b). However, an almost verbatim statement without the name change also appeared on the University of Minnesota Press website in December 2007 (e.g., Ben-Porath and Tellegen 2007d). Like us, other psychologists downloading these statements at different times or from the publisher’s or the distributor’s sites at first may miss these notable changes. Furthermore, simply changing the scale’s name and not addressing the range of concerns in Judge Bergmann’s ruling (see Butcher et al. 2008, p. 206) is unlikely to change the scale’s fundamental use or prejudicial application.

Such a significant and highly unusual name change after 17 years of a scale’s use and without any change to its item content or scoring is unprecedented in the psychological literature. The FBS name change contrasts sharply with the convention of supplementing the names and abbreviations given to the standard scales developed by Hathaway and his colleagues in the 1940s–1950s with scale numbers (e.g., Hypochondriasis (Hs), scale 1; Depression (D), scale 2; Hysteria (Hy), scale 3; and so forth). Although some MMPI experts prefer using scale numbers, which is especially useful shorthand for code types (e.g., 12/21, 123, 49/94), the original names are still used in contemporary MMPI-2 and MMPI-A research and texts. Furthermore, the FBS name change occurred immediately after a court ruling that its name was “prejudicial”, without any accompanying

² We were not informed of the journal’s policy for disclosure of potential conflicts of interest at the time of the publication of Butcher et al. (2008). The disclosure footnote in this response is applicable to that publication as well.

notice to psychologists or cautions or any recommended differences in interpretation.

Our primary concern is that a change in name is not equivalent to a change in clinical application. To reiterate, the existing body of empirical data, when carefully considered, does not provide adequate support for the validity of the FBS as a tool for assessing malingering. A secondary concern points to the need for a manual or manual supplement to assist psychologists when a new measure like the FBS is introduced to an assessment standard like the MMPI-2, rather than reliance on web site pages that can be altered without any indications that revisions have been made to the statements.

Concerns About Potential for Gender Bias Remain

Almost half of the 43 items on the FBS, when scored in the deviant direction, produce a differential responding between men and women of 5% or higher, with women more likely to respond in the deviant direction than men. Only one FBS item produces a similar difference in endorsement frequencies in which men are more likely to respond in the scored direction. Not surprisingly then, women produce higher scores than men on the FBS. As Ben-Porath et al. (2009, p. 76) admit, this gender effect was recognized early in the scale's history and "led to adjustments in the recommended raw score cutoffs for FBS (24 men; 26 women; Lees-Haley 1992)". Yet, Ben-Porath and Tellegen (2007a, b) have more recently recommended the same raw score cutoffs for men and women on the FBS (i.e., above 22 and above 28). As we pointed out in our article, the practical outcome of this recommendation is that the interpretive statement that an individual's FBS score raises "very significant concerns about the validity of self-reported symptoms" (Ben-Porath and Tellegen 2007a, b) occurs at a T score equivalent of 87 for women, but 95 for men, almost a full standard deviation lower for women, thus lowering the threshold for women to be identified as potentially malingering. Ben-Porath et al. (2009) ignored this substantive concern.

In response, Ben-Porath and colleagues present gender effect sizes from the MMPI-2 normative sample (Cohen's $d=0.53$), a general clinical sample ($d=0.59$), a "mild traumatic brain injury (TBI)" sample ($d=0.39$), and a chronic pain sample ($d=0.14$) in Table 6 of their response. Two of the clinical samples (i.e., the general clinical and mild TBI samples) in Table 6 (Ben-Porath et al. 2009, p. 76) are cited as being from "Greve et al. (2006a, b)". The chronic pain sample in Table 6 has no citation, so it is unclear if that sample is from a published source. Ben-Porath et al. (2009) identified the effect sizes for the normative sample and general clinical sample as "moderate" and the chronic pain and TBI samples as "small" (Cohen (1988, pp. 25–26) offered the following "conventional

operational definitions" for effect sizes: small=0.20, moderate=0.50, large=0.80). Next, they performed chi-square analyses on the samples of TBI and chronic pain patients whom they also classified into malingering and nonmalingering groups and concluded that there were no gender differences in the false positive error rates in these two samples.

However, Ben-Porath et al. (2009) did not provide sufficient information about the classification methods used to assign patients to malingering and nonmalingering conditions for their analyses presented in Tables 7–8 or Fig. 4 of their response to our concerns about possible gender bias. This omission does not allow the reader to determine if the concerns we raised about how criterion groups have been defined in previous research on the FBS are relevant to these new analyses as well (see Butcher et al. 2008, pp. 195–197). The sample sizes for the pain patients in Tables 6 ($n=301$) and 8 ($n=476$) reported by Ben-Porath et al. suggest that two different chronic pain data sets were used for the analyses presented in their response. The sources for these samples were not given. Without basic information about these samples and classification methods, the analyses in their response do not sufficiently respond to the concerns that the FBS may be biased against women.

Furthermore, findings of a moderate gender effect in the MMPI-2 normative sample and the general clinical sample provided in Table 6 of their response is a clear indication of a potential for bias when using the same raw score cutoffs for men and women or nongendered T scores as in the case of the MMPI-2-RF (see Butcher and Williams 2009 for a discussion of nongendered T scores). Others as well report substantial gender effects for the FBS (e.g., Dean et al. 2008), and the MMPI-2-RF manual demonstrates gender differences in the correlations between the MMPI-2-RF version of the FBS and Restructured Clinical Scale 1 (RC1; a measure of somatization). For men, the correlation between FBS-R and RC1 is 0.58 or 34% of the variance; yet the correlation for women is 0.66 or 44% of the variance (Ben-Porath and Tellegen 2008; p. 38, 42). Our concern remains: Use of the FBS as currently recommended constitutes inherent bias against women because it more often classifies women as malingering as compared to men. There is no theoretical or empirical basis for the FBS to indicate that women are more likely to falsify a psychological evaluation than are men, which raises questions about the construct validity of this measure.

Transparency vs. Nontransparency in MMPI-2 Decisions

Many of the comments in Ben-Porath and colleagues' (2009, pp. 78–79) section about our article's critique of the

decision to add the FBS to the MMPI-2 represent the logical fallacy known as *appeal to emotion*, or an attempt to win an argument by producing strong emotions in place of evidence for a claim. Ben-Porath and colleagues' (2009, pp. 79) argument is based on highly charged assertions:

“ . . . 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?”

In fact, Ben-Porath et al. (2009, p. 79) explicitly state that they will not deal with the substance of our comments about the publisher's decision-making process or the guidelines for use of the FBS (i.e., Ben-Porath and Tellegen 2007a, b) posted on the Internet:

“Because we do not wish to reinforce this conduct, we will not respond to the specifics of Butcher et al.'s analysis of the reviews.”

Substance is important, as is correcting the misrepresentations by Ben-Porath et al. (2009) of the appropriateness of including a description of our concerns about the review process in our article. A plaintiff's attorney who was part of the team in *Williams v CSX* obtained the documents in question through a request to the University of Minnesota under the Minnesota Data Practices Act. The University of Minnesota's Records and Information Management Office in the General Counsel's Office determined the information in them was not private and thus released the documents. Prior to submitting Butcher et al. (2008) for publication, we took the precaution of verifying with the University's General Counsel that the documents we planned to cite in our article were part of the public record and that we were free to distribute or quote from them as we saw fit.

Subsequently, the documents were included as evidence in *Williams* and were used in the cross-examination of the defense's expert witness (i.e., Ben-Porath). The plaintiff's expert (i.e., Butcher) was asked to review and comment on them as part of his services and was asked questions about them during his testimony. The same documents have been used in other cases as well. It is not unreasonable to assume

that other expert witnesses may be asked their opinions about the documents in other cases involving the FBS. Therefore, it seemed important to us to bring the debate about the review process, as captured in the documents, to a broader psychological audience, to assist our profession in assessing the merits of FBS inclusion in standard MMPI-2 score reports.

Like all scientific endeavors, the peer review process should be transparent and subject to review and comment, especially in cases when it is used to justify highly controversial decisions that impact people's lives. In our excerpts from the FBS reviews in Butcher et al. (2008), we did not identify the reviewers by name. We included direct quotes from each reviewer to illustrate the lack of agreement among these experts about an appropriate cutoff score for making FBS interpretations, a key point in our critique (we used direct quotes, as opposed to paraphrases, to limit the introduction of bias). Given the controversy surrounding the use of the FBS, we felt it important to provide access to the actual documents if professionals have further questions about our interpretation of what is in them.

What is more damaging to the scientific review process: transparency or secrecy? Are these documents and their use in forensic cases relevant to psychologists testifying for either the defense or plaintiffs? And, given that our conclusions about the reviewers' comments differ so significantly from the recommendations made by Ben-Porath and Tellegen (2007a, b), should other psychologists be made aware of the controversy and provided information to reach their own conclusions? These were the questions we considered when deciding whether to include our critique of the review process in Butcher et al. (2008). Although scientific journals routinely keep the identity of their reviewers confidential and some even allow for anonymity for the authors of the article under review, this does not mean that the underlying methodology of a review process, especially when consensus among the reviewers is claimed as justification for a change in clinical practice, is reasonably withheld from scrutiny by the field.

Distortions About and Distractions to Our Criticisms of FBS

Ben-Porath et al. (2009) asserted that we engaged in a selective and distorted review of the research literature in reaching conclusions about the FBS. Two points are in order: First, our goal of addressing potential biases in using the FBS did not necessitate providing the reader with an exhaustive literature review. The issue at hand cannot be resolved by comparing the number of accumulated asterisks

for and against the FBS. Rather, our legitimate and focused concern was with reporting some of the studies that raise significant questions about FBS use as a supplemental tool for assessing malingering. The existence of such studies in and of itself raises doubt about the use of the FBS. Second, we pointed out examples of biases and errors in some of the most widely cited studies (including meta-analytic) used to support the FBS. A scale like the FBS would be an invaluable clinical tool if and only if it had adequate empirical support for measuring malingering. However, we have too many concerns to support its use for clinical decisions based on the extant body of FBS research that we detailed in Butcher et al. (2008).

In addition to suggesting that we disregarded the literature on malingering, Ben-Porath et al. (2009) allege we misunderstood the conceptual foundation of the F Scale and misattributed statements to Hathaway and McKinley's (1942) original MMPI manual. These are examples of underlying logical fallacies involving distortions of our original comments (i.e., *straw person* fallacies) or distractions from our key points (i.e., *red herring* arguments). In the next several sections, we respond to those arguments.

Reply to the Distortions of Our Description of the F Scale Development

Ben-Porath and colleagues (2009) assert that the following description we provided in our article (Butcher et al. 2008, p. 198) about the development of the F Scale is erroneous and misattributed to Hathaway and McKinley (1942):

“Next, only items endorsed infrequently in the original Minnesota normative sample (i.e., no more than 10% of the sample endorsed the item in the scored direction) were included on the F scale, based on the premise that only individuals trying to exaggerate or malingering psychopathology will endorse items from broad and inconsistent problem areas that are in excess of what most patients would endorse and do not represent actual syndromes or disorders (Butcher and Williams 2000).”

They go on to assert that Butcher and Williams (2000) are also in error in their description of the development of the F Scale. They point out that the term “over-reporting” is not in the original manual and Hathaway and McKinley (1942, p. 9) suggest that carelessness or poor comprehension are the only “known interpretation” for “a high F score”.

However, had Ben-Porath et al. (2009) read two pages later in the original manual, they would have found this discussion of the development of the F Scale:

“The F score (Table IV) is derived from a group of 64 items that have been very infrequently answered in the

scored direction by normal persons. All the items are answered in the infrequent direction less than 10 per cent of the time by normals, and the percentage is but little higher for miscellaneous abnormal subjects. Very few of the items are intercorrelated to a significant extent. *Therefore, these items as a group do not form a scale in the usual sense but merely indicate whether or not the subject has made many responses that are avoided by most persons.* In reality if the items are examined it will be seen that a high score would not indicate any known pattern of symptoms.” (Hathaway and McKinley 1942, p. 11, italics added).

“Whether or not the subject has made many responses that are avoided by most persons” is, in essence, synonymous with the concept of overreporting, a term introduced later. Furthermore, Hathaway in the early days of the development of the MMPI (see also Dahlstrom and Welsh 1956) provided extensive information about F that is consistent with the information we provided in Butcher et al. (2008) and Butcher and Williams (2000). At the risk of belaboring the point, consider the following from Meehl and Hathaway (1946, p. 537):

“It was, of course, immediately possible to consider the F score as an evidence of this attempt to malingering and obtain fallaciously bad scores on other scales . . . From this experiment it appeared that F was a very good device for identifying the intentional faking that could be set up in an experimental situation.”

Ben-Porath et al. (2009) spent extensive time suggesting our confidence in the well-validated MMPI-2 validity scales, most notably F, is misguided and reflects a “double standard” that disadvantages the FBS. However, our discussion of its empirical development and subsequent validation was appropriate and accurate. The F Scale was derived as a means of empirically examining the tendency of some individuals to endorse items that are rarely endorsed in the general population. The F Scale highlights infrequent item endorsement and suggests several potential reasons for rare response endorsement such as random responding, endorsement of unusual symptoms, inattention to content resulting from such problems like reading or comprehension difficulties, and symptom exaggeration.

Reactions to Comments About Item Overlap

In response to our concerns that the rationally selected and not empirically validated items for FBS had considerable overlap with items on scales measuring somatization and somatic symptoms, as well as scales measuring defensiveness, Ben-Porath et al. (2009, p. 67) countered that we “set up unrealistic expectations and selectively applied them to

FBS.” They pointed out that of the 60 MMPI-2 F items, 40% appear on one or more of the scales related to thought disorders (e.g., Clinical Scale 6 or Pa and Scale 8 or Sc, Content Scale Bizarre Mentation or BIZ).

Yet, by computing the proportion of the 24 items that appear on both the F Scale and any of the three other scales (i.e., Scales 6, 8, and BIZ) with the total number of F items (i.e., 60), they have substantially exaggerated the overlap of F items on the Clinical Scales 6 and 8. The actual overlap between F and Scale 6 is 23% (i.e., nine F items appear on the 40 item Pa Scale). The actual overlap between Scale F and Scale 8 is 19% (i.e., 15 F items appear on the 78 item Sc Scale). Only F and the Content Scale BIZ can be characterized by their 40% estimate of item overlap: BIZ’s overlap with F is 43% (i.e., ten F items appear on the 23 item BIZ Scale). A comparison of the 24 items across all three scales that overlap with F with the total number of items on the three scales (i.e., 141) indicates that only 17% of the items on those scales are made up of F items.

Moreover, the key difference between the F Scale and the FBS is that the item overlap between F and the Clinical Scales related to thought disorders resulted from actual empirical rarity of the items in the general population and the occurrence of symptoms in the patient populations. In empirical scale development, item overlap between scales is not necessarily problematic in that two measures may actually be focusing upon different but related constructs. Item overlap can simply reflect the fact that one measure bears some relationship to the other as in the case of individuals with psychotic disorders endorsing rare symptoms (e.g., hallucinations, delusions). Item overlap across measures becomes problematic when the content results from a faulty assumption such as that contained in the FBS development. One professional selected the FBS items without any empirical verification that the FBS items can differentiate malingerers from other individuals with somatoform disorder, posttraumatic stress disorder, or somatic problems, and as we pointed out in our article, the FBS is highly correlated with empirically validated scales measuring somatoform disorders and somatic problems. Thus, the overlap between the FBS and scales measuring somatic symptoms remains problematic because the constructs of malingering and actual somatic symptoms are not related, but in fact, quite divergent.

Ben-Porath and colleagues’ (2009, p. 66–67) analysis showing high correlations between F and Clinical Scale 8 is another distraction from concerns about the FBS as it simply confirms what has been known of that relationship and widely reported, since the 1940s (e.g., Dahlstrom et al. 1972; Nichols and Crowhurst 2006). Meehl and Hathaway (1946 pp. 535–536) illustrate:

“In addition, however, it was early discovered that schizoid subjects and subjects who apparently wished

to put themselves in a bad light also obtained high scores. The schizoid group obtained high scores because, owing to delusional or other aberrant mental states, they said very unusual things in responding to the items and thus obtained high F scores.”

There is construct overlap between F and Scales 6, 8, and BIZ (i.e., psychotic symptoms are rarely endorsed items by subjects in the general population), and interpretations of F have taken that empirically demonstrated finding into consideration since the 1940s. In contrast, no empirically based rationale has been provided for the item overlap among the rationally selected FBS items and MMPI-2 scales measuring somatization, somatic symptoms, and defensiveness.

Response to the Distortions About Use of Standardized T Scores

Ben-Porath et al. (2009) correctly described the problems with the original T scores developed by Hathaway and McKinley (1942). However, this is a distraction because problems with the original T scores, despite Ben-Porath and colleagues’ (2009) contrary assertions, are not incongruous with the fact that all MMPI-2 validity scales, with the exception of FBS, are interpreted based on separate T scores for men and women. Because Lees-Haley et al. (1991) did not seek access to the MMPI-2 normative sample when they developed the FBS, he and others used raw scores for their interpretive recommendations. Yet, T scores for the FBS were developed and published by Greene (2000), but, again, unlike all other MMPI-2 validity scales, T scores were not part of the guidelines for interpretations of this scale when it was incorporated into the Extended Score Report (Ben-Porath and Tellegen 2007a, b).

The use of the same raw score cutoffs for both men and women instead of gender-based norms are related to our concerns about the inherent gender bias in the items selected for the FBS. For example, very early in the history of the MMPI, Hathaway and McKinley (1940) noted that women endorsed more items on the Depression Scale than men. They were concerned that there may be a general response style difference in men’s and women’s responding to MMPI items that are unrelated to the construct being measured by a given scale. For that reason, they developed gender-based norms for the MMPI where a woman’s response was compared with other women and a man’s response to other men. There is evidence reported in our original article, as well as in the response of Ben-Porath et al. (2009; see section above on gender bias) to our article that the potential for gender bias in use of the FBS remains a serious issue with the potential to harm women.

Reply to Distractions About the Variable Yardstick for FBS Cutoffs

Potential bias in using the FBS is partially based on the simple fact that there is no clear consensus of opinion regarding the proper cutoff score for presumptive evidence of malingering. Recommended cutoff scores for the FBS vary greatly. Widely varying recommendations are not just historic with changes coming about as new evidence accumulates, as is the case with other MMPI-2 scales. The cutoff recommendations even among the reviewers the University of Minnesota Press consulted when deciding to add the FBS to the MMPI-2 differed significantly (see Butcher et al. 2008, pp. 204–206).

Not surprisingly, even the best informed clinician will have wide latitude in selecting an FBS cutoff score, depending on which research studies and experts he or she gives credence to. Unless the publisher's current guidelines of cutoff raw scores greater than 22 and greater than 28 (Ben-Porath and Tellegen 2007a, b, c, d) capture a previously unknown truth about the FBS and supersede all previous recommendations—including those made by the publisher's reviewers within the past year (Butcher et al. 2008, pp. 204–206)—it is difficult to conceive how such wide-ranging FBS cutoff scores can be ascertained as being right or wrong. The cutoffs in the web guidelines (Ben-Porath and Tellegen 2007a, b, c, d) do not reflect a clear consensus among even the publisher's own assessment experts.

The potential bias of using the FBS with brain-injured individuals was noted by Greve et al. (2006b, p. 503), who concluded that their obtained data “are consistent with recent findings that elevations (on FBS) can occur above standard cutoffs in patients with significant neurological injury (Greiffenstein et al. 2002).” We concur with their conclusion. However, Greve et al. (2006b, p. 503) go further to allege that such findings might actually represent “a false negative that was outside the scope of the Slick et al. (1999) criteria”. Given this ambiguity one must ask if the FBS proponents are willing to relax their own criteria for interpreting FBS scores, how can we reasonably expect clinicians to rely on those criteria? FBS cutoff scores appear to function in a manner analogous to floating anchors.

In response to our criticism that lower FBS scores are common in more severe TBI cases partially due to anosognosia (lack of awareness of physical disability), the response of Ben-Porath et al. (2009) is based on a false assumption that Glasgow Coma Scale (GCS) results of Greve et al. (2006b), which were obtained immediately after the head trauma incident, are indicative of brain injury severity long (e.g., 21 months) after recovery. GCS scores provide a crude but useful gauge of injury severity proximal to the time of injury, but not beyond the recovery

period months or years later. The MMPI-2, FBS, F, F_b, and F_p were administered long after the GCS and long after recovery from the injury. Countless intervening variables could occur that would have an impact on MMPI-2 scores that would dwarf that of long past GCS scores. Had neuropsychological test results been obtained at the time of MMPI-2 administration, these would have provided an appropriate measure of brain injury severity.

Research findings contradict the assertion of Ben-Porath et al. (2009) that lack of deficit awareness (anosognosia) has no impact on MMPI-2 scores in more severely brain-injured patients. The association of scores on the L scale with degree of cerebral impairment is well established in the MMPI-neuropsychological literature (Dikmen and Reitan 1974, 1977; Gass 2006; Gass and Ansley 1994; Gass et al. 1999). However, the paradoxical severity effect in TBI is only partly related to the impaired awareness in more severely injured individuals. A significant body of empirical literature suggests that the most powerful contributor to elevated MMPI-2 scores in mild TBI is somatoform symptomatology (Greiffenstein and Baker 2001; Putnam and Millis 1994; Youngjohn et al. 1997). Somatoform symptoms, highly represented on the Hs, Hy, and Health Concerns scales, not surprisingly provide a major contribution to scores on FBS.

Reiteration About the Lack of Appropriate Controls in FBS Studies

The existing empirical literature regarding FBS paints a mixed picture of the scale's utility in assessing malingering. A fundamental problem of potential bias in using the FBS is based on the fact that if an individual scores high on FBS, it is unclear what portion of the elevation is attributable to the individual's report of physical, psychological, and/or other motivational issues. The majority of studies fail to adequately control for emotional status and psychological diagnosis. This is important considering recent evidence that several MMPI-2 clinical scales alone account for 66% of the variance in FBS scores (Downing et al. 2008). FBS scores might actually be higher in somatization disorder than in cases involving conscious faking of symptoms. Guez et al. (2005) found very high FBS scores in a sample of chronic neck pain patients who were not seeking compensation, passed a malingering test, and performed within normal limits on a neuropsychological test battery. Their data indicated that the high FBS scores in this clinical sample were a reflection, not of malingering, but of “somatization and inadequate coping” (abstract, p. 151). The fact that this sample was incorporated into the Nelson et al. (2006) meta-analytic study as a malingering sample clearly exemplifies the problem of method bias that characterizes some of the more widely cited FBS research.

Studies often employ mild TBI samples consisting of individuals manifesting a late postconcussive syndrome. A substantial body of research suggests that, in many cases, their persisting complaints represent symptoms of a somatization disorder (McCrae 2008; Youngjohn et al. 1997). It is difficult to reject the longstanding recognition that unconscious psychological factors can cause people to experience (and report) an unusually large number of physical symptoms and preoccupations. Nobody disputes the fact that financial incentives are significantly related to levels of symptom reporting. The nature of this relationship has not been fully explored. What is in dispute is the notion that there is a validated empirical basis for asserting that any FBS cutoff score reliably differentiates between malingering and somatization. In the literature review of Nelson et al. (2006), the largest effect size on FBS was, by far, attributable to a sample of nonmalingerers who appeared to have somatization disorder (Guez et al. 2005).

Somatization and other psychological factors are a potential confounding variable in the Greve et al. (2006b) study as well as in the vast majority of other investigations involving so-called “known” groups in which psychiatric diagnoses are absent or unreported.³ As we pointed out in our article, FBS studies, including ones cited by Ben-Porath et al. (2009) in their section titled “Concerns about False Positives”, use being in litigation as a proxy for having a “known incentive” to malingering without any external validation indicating how many subjects in any given litigant group are actually malingering symptoms.

Is it Likely that Our Eating Disorder Sample Was Contaminated with Malingerers?

We reported that 8% of a sample of 2,054 women with eating disorders in voluntary inpatient treatment reached a score of 30+ on the FBS (Butcher et al. 2008). As we described, these women had otherwise valid MMPI-2 profiles according to accepted criteria for Cannot Say, Variable Response Inconsistency (VRIN), True Response Inconsistency (TRIN), F, L, and K. Given the extensive and documented psychological and medical problems in this patient sample, we expressed valid concerns about false positives. Our concerns are heightened in light of the publisher’s guidelines for use of the FBS (Ben-Porath and Tellegen 2007a, b) that refer test users to a chapter by Greiffenstein et al. (2007). This chapter authoritatively states that scores of 30+ can be used to identify malingering with “the greatest confidence irrespective of gender,

medical, or psychiatric context” (p. 229) since “scores of 30+ never or rarely produce false-positive errors” (p. 228).

Ben-Porath et al. (2009, p. 78) responded to our data with the assertion that “a high proportion of eating disorders patients may have a disability claim and therefore financial incentive” to malingering, which would be reflected in elevated FBS scores, and we disregarded this possibility when reporting our findings. This suggestion is both speculative and inaccurate in that the eating disorder sample we included in our article consisted of clinical patients who had undergone extensive professional evaluation and were not seeking payments for disability. In addition, it seems illogical that a woman would voluntarily seek treatment for an eating disorder at the same time that she is seeking payment for having the disorder. We will acknowledge that there may be a remote possibility that some eating disorders patients in this voluntary treatment setting could be seeking disability payments, and somehow this escaped the notice of the evaluation and treatment team. However, given our large sample size and the improbability of a significant number of such occurrences, it is highly unlikely that there would be sufficient cases with such incentive to malingering in our inpatient sample to have a demonstrable effect on the percentage of women with elevated FBS scores of 30+ reported in our article (Butcher et al. 2008). Guidelines that suggest no false positives regardless of setting for scores of 30+ are belied by the data from this clinical setting of very ill women with otherwise valid MMPI-2 profiles.

An anonymous reviewer of this article asked for a comparison of the 8% false positive rate for FBS >29 with false positive rates in this eating disorder inpatient sample for F, F_b, and F_p. The 2,054 subjects used to calculate the 8% rate for FBS produced valid MMPI-2 profiles based on standard validity criteria recommended for research studies (i.e., subjects were eliminated based on CS, VRIN, TRIN, F, L, and K as detailed on page 203 of Butcher et al. 2008). Therefore, in this sample, the elevations for the F family were as follows:

$$\begin{aligned} F &= 0\% \\ F_b &= 9\% \\ F_p &= 0.2\% \end{aligned}$$

These rates directly compare with the 8% rate for FBS scores greater than 29, requested by the reviewer. Table 2 of Butcher et al. (2008, p. 203) included seven other cutoff scores for the FBS for this sample of eating disorder inpatients. These rates ranged from 11% for the publisher’s recommended cutoff score of greater than 28 to 62% for the original cutoff recommendation of 20.

If we examine the entire sample of 2,273 eating disorders patients and do not eliminate any subjects with invalid MMPI-2 profiles based on F, but do eliminate subjects using the other standard validity criteria recom-

³ The term “known-groups design” is a misnomer, since it is typically based on probabilistic and circumstantial evidence, incomplete information, and rarely on direct knowledge.

mended for research studies described in Butcher et al. (2008; subjects eliminated based on CS, VRIN, TRIN, L, and K), 4% are elevated on F. Using these validity criteria, the comparable rates for FBS are 9% (FBS raw scores greater 29) and 12% (raw scores greater than or equal to 29—the publisher’s current recommended cutoff score for profile invalidity). These rates are based on a total sample of 2,146 inpatient women.

It is important to consider these rates in the context of how these various measures of overreporting were developed and how they are currently used (see “Reply to the Distortions of Our Description of the F Scale Development” section). F, F_b, and F_p were all empirically validated to include items rarely endorsed in normative settings and, in the case of F_p, in psychiatric settings. There was no empirical validation for the FBS items. Interpretation of elevations on the F family of scales is limited to performance on the MMPI-2, a measure of personality and psychopathology. Current claims by FBS proponents suggest that it “can be helpful in cases where someone with a mild or non-existent brain injury is trying to appear seriously dysfunctional or disabled but not psychotic” (Lees-Haley as quoted in Pearson Assessments 2007a, b).

The potential harm to patients from false positives when the using the F family of scales is much lower in our eating disorders sample, given the rates presented above and the more conservative interpretive guidelines for these scales. When MMPI-2 profiles with elevations on F greater than 100 are included in the sample ($N=2,146$), over three times as many women receiving inpatient treatment for eating disorders (i.e., 12%) are identified as having “very significant concerns about the validity of their self-reported symptoms” using the publisher’s recommended cutoff score for the FBS. This compares to 4% for elevated F scores. When the established validity criteria for the MMPI-2 are used to eliminate invalid protocols from the sample ($N=2,054$), 55 times as many women being treated for eating disorders (i.e., 11%) are identified by the FBS in contrast with F_p (i.e., 0.2%). The F_b is a measure of performance on the second half of the MMPI-2 booklet. The MMPI-2 standard scales, whose items are contained in the first half of the booklet, can be interpreted even in the presence of an elevated F_b. The individual’s self-report on those key MMPI-2 scales, as well as their reports of symptoms related to brain injuries, are not challenged on the basis of an elevated F_b (see Butcher et al. 2008, pp. 197–198 for further discussion of comparisons of the FBS with the F family of validity scales).

Examples of Forensic Cases Involving the FBS

Ben-Porath et al. (2009, p. 79) erroneously characterized our article as appearing “to be advancing a legal

argument” and indicated, “Their selective use of legal authority would not be accepted in a court of law.” Even a cursory read of our article reveals that we made no representations in it about legal arguments, legal authority, or legal analyses, despite their contrary assertions. We did include a four-paragraph summary highlighting one of three *Frye* challenges in Florida courts about the use of the FBS in expert witness testimony and included a direct quote with the judge’s conclusions about FBS bias and subjectivity.

Although Ben-Porath et al. (2009) acknowledged that an American Bar Association (ABA) rule does not apply to a journal article written by psychologists, they nevertheless made the following *ad hominem* assertions that we “. . . would be subject to sanctions . . .” if we were attorneys and “. . . made such unbalanced representations to a court,” citing out of context for the profession of psychology ABA “Rule 3.3(a)(2) Candor Toward the Tribunal” (Ben-Porath et al. 2009, p. 79). They go on with another *ad hominem* claim that one of us (i.e., Butcher) inappropriately swayed Judge Bergmann in *Williams v CSX Transportation* with:

“. . . testimony that is inconsistent with the scientific literature and characterized by many of the same flaws we’ve 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” (Ben-Porath et al. 2009, p. 82).

Judge Bergmann weighed all the evidence presented before him and reached a conclusion that the FBS was too subjective to be included as part of expert witness testimony in his courtroom. Butcher testified truthfully for the plaintiff about his conclusions based on the psychological literature regarding the use of the FBS, contrary to assertions by Ben-Porath et al. (2009) that his testimony was erroneous, misleading, and inflammatory.

Ben-Porath et al. (2009) failed to disclose in their descriptions about this case that the defense in *Williams v CSX Transportation* had an extensive array of eight experts, including Ben-Porath who testified as a defense expert at the *Frye* Hearing. Lees-Haley, the author of the FBS, was among the several defense experts who provided affidavits in support of the utility of the FBS. The judge did not find the evidence presented by this defense team as persuasive as the plaintiff’s, and this case went to trial, without the FBS characterization of the plaintiff as malingering. The jury found in favor of the plaintiff.

Since we wrote Butcher et al. (2008), two new judges in *Frye* hearings in Florida and one in California have ruled against including the FBS as part of expert witness

testimony (note that the Florida cases are not isolated to Florida's 13th Circuit as asserted by Ben-Porath et al. 2009). Judge Hoy's ruling in *Stith & Stith v State Farm Insurance* (2008, p. 2) further illustrates the concerns over using the FBS in court testimony:

"The evidence presented at the hearing supports the conclusion that the FBS is not an objective measurement of malingering, exaggerating or over reporting of symptoms. The FBS is inherently unreliable because it scores points in malingering, exaggerating or over reporting when a patient has true symptoms of physical injury or physical distress. The FBS has the significant potential to negatively impact persons with true disabilities. The evidence presented showed that the test is biased against women because they tend to score higher on the FBS than men, particularly when they have verifiable injuries."

Particularly revealing of recent court's deliberations and opinions about the FBS is the transcript of proceedings before Judge Winesett of Florida's 20th Judicial Circuit in a *Frye* hearing (*Limbaugh-Kirker et al. 2009*). Judge Winesett detailed the evidence she used in arriving at the decision to exclude:

". . . any testimony of Dr. Larrabee using reference to the Fake Bad Scale as a scientific means of assessing malingering, exaggeration, or over-reporting of the Plaintiff or any reference to it to bolster his opinion that Plaintiff is malingering, exaggerating, or over-reporting, or not truthful or credible" (p. 12).

According to Judge Winesett, she considered the following before arriving at her decision (*Limbaugh-Kirker et al. 2009*):

- Trial and deposition testimony of Dr. Glenn Larrabee (an author of numerous articles in support of use of the FBS)
- Deposition testimony of Dr. Manfred Greiffenstein (an author of numerous articles in support of use of the FBS, including the book chapter recommended on the publisher's website—Greiffenstein et al. 2007)
- The article by Butcher et al. (2003)
- The article in the *Wall Street Journal* (Armstrong 2008)
- Four Florida trial court opinions:
 - *Vandergracht & Vandergracht v Progressive Express, USAA Insurance Company, & TIG Insurance Company, 2005*
 - *Davidson v Strawberry Petroleum, Inc. & Haddle, 2007*
 - *Williams v CSX Transportation, Inc., 2007*
 - *Stith & Stith v State Farm Mutual Insurance, 2008*
- Cases submitted by the defense, which the judge described as follows: "Those cases mainly related to

the requirements of *Frye* and instances in which *Frye* hearings were held with respect to particular matters" (*Limbaugh-Kirker et al. 2009*)

- Oral arguments of plaintiff and defense

The above list is inclusive of all that Judge Winesett described as informing her decision. It is important to note that Butcher had no role in the case heard by Judge Winesett. Indeed, he only learned of *Limbaugh-Kirker & Kirker v Discosta* after the judge made her decision. Although the article he co-authored with Arbisi, Atlis, and McNulty in 2003 was among the materials considered by Judge Winesett, she indicated in her ruling that Butcher et al. (2003) was "pointed out in the testimony of Dr. Greiffenstein and highly criticized by him" (*Limbaugh-Kirker et al. 2009, p. 9*), and contrary to the discounting by Ben-Porath et al. (2009) of Judge Bergmann's decision making and ruling in the *Williams v CSX Transportation* case and the two other Hillsborough County cases (i.e., *Vandergracht* and *Davidson*), Judge Winesett indicated, "I am persuaded by the reasoning and conclusions of those courts, particularly that of Judge Bergmann because he did a very detailed ruling" (p. 11).

Unfortunately, in their other attempts to downplay the challenges to the use of the FBS in forensic cases, Ben-Porath et al. (2009, p. 79) provided the following inaccurate information:

"Numerous board-certified clinical neuropsychologist experts report admission of FBS testimony into evidence with some testifying that they have never had FBS testimony excluded (e.g., *Upchurch v Broward Co. School Bd. 2008*;⁴ *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."

Both cases described in this excerpt are from Florida, which makes the following comment from Judge Winesett relevant (*Limbaugh-Kirker et al. 2009, pp. 10–11*):

"The Court also notes that notwithstanding the fact that this test was designed in 1991, there is apparently no reported Florida state court cases in which evidence regarding the Fake Bad Scale Test has been allowed."

A closer examination of facts in the two cases cited above by Ben-Porath and colleagues (i.e., *Solomon* and

⁴ *Upchurch v Broward Co. School Bd 2008* is cited in the reference list as "*Upchurch v School Board of Broward Co. (2009)*."

Upchurch) reveal that they do not support the use of the FBS in forensic settings as indicated by Ben-Porath et al. (2009). For example, the plaintiff's attorney in *Solomon & Solomon v TK Power & Goodwin* (2008) voluntarily withdrew her *Frye* objection to the FBS after the deposition of Paul Kauffman, one of the coauthors of Ben-Porath et al. (2009). Diane Weaver of Harrell and Harrell Law Firm (personal communication, January 6, 2008) reported that she decided as a trial tactic after taking his deposition that her client would likely receive a very substantial award if the jury heard his testimony about the FBS. Following her announcement that she was withdrawing her objection to the testimony about her client's FBS score, the defendant offered additional money to settle the case. The case settled.

Similarly, Ben-Porath and colleagues cited *Upchurch v. Broward Co. School Bd.*, 2009 in the same context as the *Solomon* case. The plaintiff in this case was not seeking payment for damages, only authorization for health care after her benefits had been terminated based on her FBS score. A *Frye* motion was filed and after the depositions were taken, the defense withdrew its reliance on the FBS, authorized treatment, agreed to pay attorney's fees, and agreed the plaintiff did not have to return to the expert who had relied upon the FBS.

In addition to the Florida cases, we recently learned of one in California, *Anderson et al. v. E&S International Enterprises, Inc. et al.* 2008. The defendants in this case were precluded from introducing evidence from the Fake Bad Scale (pp. 2–3):

“The court finds that the Fake Bad Scale is a “new scientific technique” within the meaning of the *Kelly/Frye* rule . . . Accordingly, as the proponent of this evidence, defendant must show that the technique is “sufficiently established to have gained general acceptance in the particular field in which it belongs” . . . Defendant has not met this burden.”

Ben-Porath et al. (2009, p. 79) also claimed that Butcher et al. (2008) “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. *United States v Bitton* 2008).” We reviewed the Memorandum Decision and Order Regarding Competency of Defendant Bitton from the US Court for the District of Utah Central Division, learning the issue before this court was whether defendant Bitton was competent to stand trial. A psychologist for the defense administered a battery of tests “in which Defendant tested at a level indicating mental retardation” (p. 2), although other testimony indicated “that according to the results of the CAST-MR test that she administered to Defendant, he

could be considered competent” (p. 3). The plaintiff requested a psychiatric exam, and the plaintiff's expert “administered a battery of tests, including the Reynolds Individual Assessment Scales (“RAIS”) test, to evaluate Defendant's competency.” The plaintiff's expert concluded “that his test results are not consistent with mental retardation but are most consistent with malingering” (pp. 3–4). The US District Judge ruled the defendant competent to stand trial.

Presenting the *United States v Bitton* for support of the use of the FBS in forensic testimony represents a *false analogy* logical fallacy. The MMPI-2 is not a measure of mental retardation, the issue at hand in this case, nor was the MMPI-2 or the FBS mentioned in the judge's ruling. Other psychological measures used to assess mental retardation were specifically mentioned. Given the nature of the evaluation and the lack of mention in the ruling, it is likely that the MMPI-2 and FBS were not part of the psychological battery relied upon by these experts to reach their conclusions. It is difficult to imagine a logical reason to conclude that this case supports the use of the FBS in forensic testimony.

Hsieh (2008) reported a jury trial in California in which the client's attorney challenged the psychologist's conclusion in cross-examination that his client was malingering based on the FBS. The psychologist acknowledged that many FBS items were symptoms that could be found in patients with chronic pain, sleep disturbances, and emotional distress. After 3 hours of deliberation, the jury returned a verdict in favor of the plaintiff (Hsieh 2008).

Hsieh (2008) described a minidebate among the plaintiff's bar regarding whether to use *Frye* or *Daubert* challenges to exclude the FBS from expert witness testimony or whether it is more effective to allow psychologists to testify about the FBS in front of the jury and be subject to cross-examination about the overlap of client's FBS item responses with their client's actual otherwise documented symptoms. As we noted above, plaintiff's attorneys may initially file *Frye* motions and then withdraw them after taking and reviewing depositions of psychologists wanting to use the FBS (see above comments about *Solomon* and *Upchurch*).

Ben-Porath et al. (2009) cited a few other cases in their response to our article, but given the short deadline for our response, we have not been able to obtain the court records to verify the accuracy of their descriptions of those cases. Given the selective and inaccurate recounting of some of their case descriptions described above, psychologists may wish to verify the accuracy and completeness of their descriptions. *Cherry picking*, or only considering incomplete case descriptions or data, is another example of a logical fallacy that can lead to inaccurate conclusions.

Concluding Comments

If the scrutinizing psychologist does a careful and objective analysis of the research supporting the FBS, we believe that he or she will recognize its questionable methodologies, its mixed results and, ultimately, its uncertain validity. One can view the FBS only in a highly favorable light by ignoring the methodological problems in the studies underlying its development or by disregarding a significant body of empirical research that casts doubt on the accuracy of the FBS, even as an adjunct to diagnosing malingering, as Ben-Porath et al. (2009) now purport it can be. We challenge the reader to resist the natural temptation to accept *carte blanche* the conclusions of FBS proponents or the current writers and, instead, independently engage in a careful review of the research literature with close consideration of methodological issues.

It is imperative that psychologists involved in psychological assessment that can have serious effects on the lives of people use instruments they fully understand. The Board of Trustees of the Society for Personality Assessment (SPA 2006) pointed out that psychological assessment typically involves a relatively brief encounter with the client and by the time a client notices that the assessor has erred, the assessment is likely to be concluded. They point out that “Psychological test reports usually become a permanent part of an individual’s medical record and are likely to follow him or her throughout his or her life, carrying with them the imprimatur of scientific fact” (p. 356). The SPA guidelines further assert that “Society as a whole is harmed both by inappropriate decisions made about individual clients as well as by the loss of confidence in professional judgment resulting from psychological assessment errors” (p. 356). We agree with the guidelines provided by SPA and conclude that, in our view, the use of a flawed psychological measure like the FBS to support conclusions that a client is “malingering” is a problematic direction for psychological assessment to take. It can only erode the confidence that the public has developed in the profession of psychology and the reputation that psychological assessment professionals have striven to develop for the past century.

Finally, we close with the following question for our colleagues to consider when evaluating the research basis of the Fake Bad Scale: Are the underlying hypotheses about the development and validation of the FBS a contemporary example of the *psychologist’s fallacy*? William James warned about this potential source of error in psychology:

“The great snare of the psychologist is the *confusion of his own standpoint with that of the mental fact* about which he is making his report. I shall hereafter call this ‘psychologist’s fallacy’ *par excellence*” (James 1890, p. 196).

References

- Anderson, M. et al. v. E & S International Enterprises, Inc et al. (2008). Case #RG05-211076, California Superior Court for the County of Alameda.
- Armstrong, D. (2008, March 5). Personality check: Malingering test roils personal-injury law. *Wall Street Journal*, p. 1.
- Ben-Porath, Y.S. & Tellegen, A. (2007a). MMPI-2 Fake Bad Scale (FBS). Retrieved September 3, 2007, from http://www.upress.umn.edu/tests/mmpi2_fbs.html.
- Ben-Porath, Y.S. & Tellegen, A. (2007b). MMPI-2 Fake Bad Scale (FBS). Retrieved September 3, 2007, from <http://www.pearsonassessments.com/resources/fbs.html>.
- Ben-Porath, Y.S. & Tellegen, A. (2007c). MMPI-2 FBS (Symptom Validity Scale). Retrieved December 4, 2007, from <http://www.pearsonassessments.com/resources/fbs.html>.
- Ben-Porath, Y.S. & Tellegen, A. (2007d). MMPI-2 Fake Bad Scale (FBS). Retrieved December 4, 2007, from http://www.upress.umn.edu/tests/mmpi2_fbs.html.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Ben-Porath, Y. S., Greve, K. W., Bianchini, K. J., & Kaufman, P. M. (2009). The MMPI-2 Symptom Validity Scale (FBS) is an empirically validated measure of over-reporting in personal injury litigants and claimants: Reply to Butcher et al. (2008). *Psychological Injury and the Law*, 2, 62–85.
- Butcher, J. N., & Williams, C. L. (2000). *Essentials of the MMPI-2 and MMPI-A clinical interpretation* (2nd ed.). Minneapolis: University of Minnesota Press.
- Butcher, J. N., & Williams, C. L. (2009). Personality assessment with the MMPI-2: Historical roots, international adaptations, and current challenges. *Applied Psychology: Health and Well-being*, 1(1), 105–135.
- Butcher, J. N., Arbisi, P. A., Atlis, M. M., & McNulty, J. L. (2003). The construct validity of the Lees-Haley Fake Bad Scale: Does this scale measure somatic malingering and feigned emotional distress? *Archives of Clinical Neuropsychology*, 18, 473–485.
- Butcher, J. N., Gass, C. S., Cumella, E., Kally, Z., & Williams, C. L. (2008). Potential for bias in MMPI-2 assessments using the Fake Bad Scale (FBS). *Psychological Injury and the Law*, 1(3), 191–209.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cumella, E. J., Wall, A. D., & Kerr-Almeida, N. (2000). MMPI-2 in the inpatient assessment of women with eating disorders. *Journal of Personality Assessment*, 75, 387–403.
- Davidson v Strawberry Petroleum & Haddle* (2007). Case #05-4320, Florida 13th Circuit in and for Hillsborough County, Florida.
- Dikmen, S., & Reitan, R. M. (1974). Minnesota Multiphasic Personality Inventory correlates of dysphasic language disturbances. *Journal of Abnormal Psychology*, 83, 675–679.
- Dikmen, S., & Reitan, R. M. (1977). MMPI correlates of adaptive ability deficits in patients with brain lesions. *Journal of Nervous and Mental Disease*, 165, 247–254.
- Dahlstrom, W. G., & Welsh, G. S. (1956). *Basic readings in the MMPI*. Minneapolis, MN: University of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook*. Minneapolis, MN: University of Minnesota Press.
- Dean, A. C., Boone, K. R., Kim, M. S., Curiel, A. R., Martin, D. J., Victor, T. L. et al. (2008). Examination of the impact of ethnicity on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Fake Bad Scale. *The Clinical Neuropsychologist*, 1, 1–7.
- Downing, S. K., Denney, R. L., Spray, B. J., Houston, C. M., & Halfaker, D. A. (2008). Examining the relationship between the

- Reconstructed Clinical Scales and the Fake Bad Scale of the MMPI-2. *The Clinical Neuropsychologist*, 22, 680–688.
- Gass, C. S. (2006). Use of the MMPI-2 in neuropsychological evaluations. In J. N. Butcher (Ed.), *MMPI-2: A practitioner's guide*. Washington, DC: American Psychological Association.
- Gass, C. S., & Ansley, J. (1994). MMPI correlates of poststroke neurobehavioral deficits. *Archives of Clinical Neuropsychology*, 9, 461–469.
- Gass, C. S., Luis, C. A., Rayls, K., & Mittenberg, W. B. (1999). Psychological status and its influences in acute traumatic brain injury: An MMPI-2 study [Abstract]. *Archives of Clinical Neuropsychology*, 14, 30.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston, MA: Allyn and Bacon.
- Greiffenstein, M. F., & Baker, W. J. (2001). Comparison of premorbid and postinjury MMPI-2 profiles in late post concussion claimants. *The Clinical Neuropsychologist*, 15, 162–170.
- Greiffenstein, M. F., Baker, W. J., Gola, T., Donders, J., & Miller, L. J. (2002). The FBS in atypical and severe closed head injury litigants. *Journal of Clinical Psychology*, 58, 1591–1600.
- Greiffenstein, M. F., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 Fake Bad Scale in detection of noncredible brain injury claims. In K. Boone (Ed.), *Detection of noncredible cognitive performance*, pp. 210–235. New York, NY: Guilford.
- 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., Brenna, A., & Heinley, M. T. (2006b). Sensitivity and specificity of the MMPI-2 validity indicators to malingered neurocognitive dysfunction in traumatic brain injury. *The Clinical Neuropsychologist*, 20, 491–512.
- 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.
- Hathaway, S. R., & McKinley, J. C. (1940). A Multiphasic Personality Schedule (Minnesota) III. The measurement of symptomatic depression. *Journal of Psychology*, 14, 73–84.
- Hathaway, S. R., & McKinley, J. C. (1942). *The Minnesota Multiphasic Personality Schedule*. Minneapolis, MN: University of Minnesota Press.
- Hsieh, S. (2008, May 19). Plaintiff turns tables on 'Fake Bad Scale'. *Lawyers USA*, pp. 1, 29.
- James, W. (1890). *Principles of psychology*, Vol. 1. New York, NY: Cosimo.
- Lamberty, G. J. (2008). *Understanding somatization in the practice of clinical neuropsychology*. New York, NY: Oxford University Press.
- Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist*, 12, 179–188.
- Larrabee, G. J. (2007). Evaluation of exaggerated health and injury symptomatology. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficit*, pp. 264–286. New York, NY: 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.
- Limbaugh-Kirker & Kirker v Dicosta (2009). *Case #06-CA-00706, Florida 20th Circuit in and for Lee County, Florida*.
- McCrae, M. A. (2008). *Mild traumatic brain injury and post concussion syndrome*. New York, NY: Oxford University Press.
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable. *Journal of Applied Psychology*, 30, 525–564.
- Nelson, N. W., Sweet, J. J., & Demakis, G. J. (2006). Meta-Analysis of the MMPI-2 Fake Bad Scale: Utility in forensic practice. *The Clinical Neuropsychologist*, 20, 39–58.
- Nichols, D. S., & Crowhurst, B. (2006). Use of the MMPI-2 in inpatient mental health settings. In J. N. Butcher (Ed.), *MMPI-2: A practitioner's guide*, pp. 195–252. Washington, DC: American Psychological Association.
- Pearson Assessments. (2007a, January 11). MMPI-2 Fake Bad Scale (FBS): FBS added to the MMPI-2 standard scoring materials [Press Release]. Retrieved September 3, 2007, from <http://www.pearsonassessments.com/news/pr011107.htm>
- Pearson Assessments. (2007b, January 11). FBS (Symptom Validity Scale) added to MMPI-2 standard scoring materials: Scale helps identify non-credible reporting [Press Release]. Retrieved November 12, 2007, from <http://www.pearsonassessments.com/news/pr011107.htm>
- Pope, K. S. (2009) Logical fallacies in psychology: 21 types. Retrieved March 31, 2009 from <http://kspeope.com/fallacies/fallacies.php>
- Putnam, S. H., & Millis, S. R. (1994). Psychosocial factors in the development and maintenance of chronic somatic and functional symptoms following mild traumatic brain injury. *Advances in Medical Psychotherapy*, 7, 1–22.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *Clinical Neuropsychologist*, 13(4), 545–561.
- Society for Personality Assessment. (2006). Standards for education and training in psychological assessment : Position of the Society for Personality Assessment. *Journal of Personality Assessment*, 87, 355–357.
- Solomon & Solomon v T. K. Power & Goodwin (2008). *Case #06-CA-00388, Florida 4th Circuit, in and for Duval County, Florida*.
- Stith & Stith v Williams & State Farm Mutual Insurance (2008). *Case #2003 CA 010945 AG, Florida 15th Circuit in and for Palm Beach County, Florida*.
- United States v Bitton (2008, July 1). Memorandum Decision and Order. Case #2:05-CR-661 TS. United States District Court of Utah Central Division.
- Upchurch v School Board of Broward Co. (2009). *OJCC Case #98-024122KSP, Florida Division of Administrative Hearings, Judges of Compensation Claims, Broward District*.
- Williams v. CSX Transportation, Inc. (2007). *Case #04-CA-008892, Florida 13th Circuit in and for Hillsborough County, Florida*.
- Vandergracht & Vandergracht v Progressive Express, USAA Insurance Company, & TIG Insurance Company (2005). *Case #02-04552, Florida 13th Circuit in and for Hillsborough County, Florida*.
- Young, G. L. (2008). Psychological injury and law: editorial introducing a journal and an association. *Psychological Injury and Law*, 1, 1–6.
- Youngjohn, J. R., Davis, D., & Wolf, I. (1997). Head injury and the MMPI-2: paradoxical severity effects and the influence of litigation. *Psychological Assessment*, 9, 177–184.