

Addressing Challenges to MMPI-2-RF-Based Testimony: Questions and Answers

Yossef S. Ben-Porath*

Department of Psychology, Kent State University, Kent, OH, USA

*Corresponding author at: Department of Psychology, Kent State University, Kent Hall, Kent, OH 44242, USA. Tel.: +1-330-672-2684; fax: +1-330-672-4826.

E-mail address: ybenpora@gmail.com (Y.S. Ben-Porath).

Accepted 20 August 2012

Abstract

Introduction of a new version of a psychological test brings with it challenges that can be accentuated by the adversarial nature of the legal process. In the case of the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF), these challenges can be addressed by becoming familiar with the rationale for and the methods used in revising the inventory, the information contained in the test manuals, and the growing peer-reviewed literature on the test. Potential challenges to MMPI-2-RF-based testimony are identified in this article and discussed in question and answer format. The questions guiding this discussion are based on the Daubert factors, established in 1993 by the US Supreme Court as criteria for gauging the scientific validity of proffered expert testimony. The answers to these questions apply more broadly to testimony in depositions, pre-trial hearings, and at trial. Consideration of the MMPI-2-RF in light of the Daubert factors indicates that the instrument has been subjected to extensive empirical testing and that a substantial peer-reviewed literature is available to guide and support its use. Information about the known and potential rate of error associated with MMPI-2-RF scores is available, and standard procedures for administration, scoring, and interpretation of the inventory are detailed in the test administration manual. Indicators of MMPI-2-RF acceptance can be cited, and criticisms of the MMPI-2-RF can be addressed with information available in the test documents and an extensive, modern, and actively growing peer-reviewed literature.

Keywords: Assessment; Forensic neuropsychology; Professional issues

Introduction

Forensic practitioners face unique challenges when a new version of a psychological test is released. An expert who uses the newer version of the test may be challenged for relying on a “new, unproven device.” On the other hand, a psychologist who uses the older version may be challenged for using an “old, antiquated instrument.” Thus, at least for a period of time, forensic users of an updated measure encounter a “damned if you do and damned if you don’t” situation that may be accentuated by the adversarial nature of the legal system.

One way to avoid this dilemma is to never update our instruments. Had this approach been adopted in the case of the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1943), forensic psychologists might still be relying on scales developed and norms collected during the late 1930s and early 1940s. Most experts and, for that matter, most legal decision-makers are unlikely to find this solution satisfactory over the long term. Alternatively, when a psychological test is revised, it is incumbent upon forensic users of the measure to (a) become familiar with the updated instrument, including the rationale for, methods used in, and outcome of the revision; (b) make an informed decision about whether to use the revised test in their forensic assessments; and (c) be prepared to defend their decision.

This article identifies challenges forensic psychologists may encounter with the introduction of the MMPI-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011; Tellegen & Ben-Porath, 2008/2011). Answers are provided to a series of questions that are likely to come up if MMPI-2-RF-based testimony is contested. In relatively rare cases, this may involve a

pre-trial effort to exclude testimony under the governing admissibility standards. More commonly, this will occur during pre-trial depositions, or at trial, under cross-examination vis-a-vis weight of the evidence considerations. Although their goals are different, the same issues are likely to be raised by attorneys seeking to exclude MMPI-2-RF-based testimony and those seeking to mitigate its impact through cross-examination. The questions and answers discussed in this article apply to both types of challenges.

The article begins with a brief overview of the revised inventory, followed by answers to a series of questions that could come up in the context of efforts to challenge MMPI-2-RF-based testimony. The framework used to articulate and address these questions is the one outlined by the United States Supreme Court in its landmark *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) decision. Although the standards outlined in *Daubert* apply to decisions regarding the admissibility of testimony in federal and now most state court proceedings, as just noted, the same or similar issues can be raised in cross-examination or deposition-taking in an effort to impact the weight a legal decision-maker, judge or jury, affords MMPI-2-RF-based testimony.

The MMPI-2-RF: a Brief Overview

A detailed account of the rationale for, methods used in, and outcome of the revision of the MMPI-2 is provided by Ben-Porath (2012). In brief, the restructuring process was initiated by Auke Tellegen to address long-known psychometric shortcomings of the original MMPI Clinical Scales (cf., Jackson, 1971; Loevinger, 1972; Meehl, 1972; Norman, 1972). The problems addressed in the revision included excessive intercorrelations between the Clinical Scales (stemming from the influence of a strong common factor and extensive item overlap) which substantially restricted their discriminant validity; and pervasive heterogeneity, including the presence of invalid subtle items, which significantly limited their convergent validity. (References to correlations between scales are more appropriately applied to scale scores rather than the scales themselves. This technical point is acknowledged here but not always applied in the paper to facilitate the flow of the text.) These problems were carried over to the MMPI-2 when the committee charged with revising the original version of the test decided to leave the Clinical Scales intact and to focus instead on collecting much-needed national norms for the inventory (Tellegen was a member of the committee that developed the MMPI-2 [Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989]. He advocated [unsuccessfully] that the revision address both the need for new norms and shortcomings of the Clinical Scales and began to work on improving the scales shortly after the MMPI-2 was published.).

Tellegen's effort, designed to address psychometric deficiencies of the Clinical Scales, culminated in the addition of a set of Restructured Clinical (RC) Scales (Tellegen et al., 2003) to the MMPI-2. Restructuring was designed to improve the Clinical Scales by removal (to the extent warranted and feasible) of a source of common variance labeled Demoralization (Tellegen and colleagues, 2006, and Ben-Porath, 2012, provide detailed discussions of this construct and how it came to artificially inflate correlations between the Clinical Scales), identification of a major distinctive (distinctive from Demoralization as well as the other identified core components of the scales) core component for each of the eight original Clinical Scales, and then development of new measures of the identified constructs. This process, guided by the current literature on associations between mood, affect, and personality, produced the nine RC Scales. The restructured scales provide substantial psychometric improvements over the original measures (see Tellegen, Ben-Porath, and Sellbom, 2009, and Ben-Porath, 2012, for recent reviews of this literature) as well as links to current models and concepts in personality and psychopathology (Ben-Porath, 2012).

At the conclusion of the monograph that introduced the RC Scales, Tellegen and colleagues (2003) noted that future restructuring efforts applied to other MMPI-2 scales could subsequently lead to the development of an improved inventory. Tellegen and Ben-Porath (2008/2011) and Ben-Porath (2012) describe the follow-up process that produced the MMPI-2-RF. Briefly, factor analyses of the RC Scales identified a recognizable three-factor higher-order structure that guided the development of three corresponding broadband measures of emotional, thought, and behavioral dysfunction. A series of analyses similar to the ones used to develop the RC Scales led to construction of 25 more narrowly focused measures of other constructs that can be validly assessed with the MMPI-2 item pool. Five revised measures of a dimensional model of personality psychopathology (the Personality Psychopathology Five [PSY-5; Harkness & McNulty, 1994; Harkness, Finn, McNulty, & Shields, 2012]) were constructed by the developers of this model. Seven revised and two new measures of protocol validity completed the 51-scale MMPI-2-RF.

Tellegen and Ben-Porath (2008/2011) provide extensive validation data for the 42 substantive scales of the MMPI-2-RF (i.e., all but the Validity Scales; external correlates of the Validity Scales are reported in the peer-reviewed literature discussed later in this article). These findings are complemented by a growing body of peer-reviewed empirical studies of various elements of the restructured inventory. This literature is described and discussed later in relation to questions regarding the empirical foundations of the MMPI-2-RF.

Potential Challenges to MMPI-2-RF-Based Testimony

MMPI-2-RF-based testimony can be challenged at different stages of litigation. As noted earlier, at the pre-trial stage opposing counsel may seek to have the testimony excluded on the basis of the standards governing admissibility of expert testimony in the jurisdiction where the trial is to take place. More often, challenges are mounted in depositions or at trial in cross-examination of an expert whose testimony rests in part on MMPI-2-RF findings. For example, if an expert concludes that a plaintiff's self-report is non-credible in part on the basis of findings on the MMPI-2-RF Validity Scales, opposing counsel may seek to have the testimony excluded altogether or may challenge it during deposition or trial testimony because of concerns about the quality of information that can be gleaned from MMPI-2-RF Validity Scale scores.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (henceforth *Daubert*), the US Supreme Court laid out a number of factors a trial judge might consider when ruling on the admissibility of proffered expert testimony. Because the issues spelled out in *Daubert* may also be raised at deposition or during cross-examination, they will serve as the framework for identifying and discussing potential challenges to MMPI-2-RF-based testimony.

The Daubert Factors

In its landmark 1993 *Daubert* decision the US Supreme Court ruled that the prior standard for the admissibility of expert testimony in federal courts, the Frye Test, had been superseded by a change in the Federal Rules of Evidence enacted by Congress in the 1970s. The Frye test was established in a 1923 federal appellate court ruling that expert testimony must be based on scientific principles or discoveries that have gained general acceptance in the particular field of the witness's expertise. At issue in the *Daubert* case was whether an expert should be allowed to testify to an opinion based on a re-analysis of published epidemiological data. Lower courts had ruled that because the analyses were new and therefore could not have gained general acceptance, the testimony should not be allowed. The Supreme Court held that the revised Federal Rules of Evidence allow for testimony based on any techniques or analyses, including novel ones, provided that they are scientifically valid. The Court charged trial judges with the gate-keeping task of determining whether challenged expert testimony meets this standard. Recognizing that in assigning this role to trial judges, it was challenging them with the formidable task of evaluating the scientific validity of proffered expert testimony, the Court included in its *Daubert* opinion some guidance about the types of factors a judge might consider in determining whether to admit challenged scientific testimony. The so-called *Daubert Factors* can be described as a series of questions as follows.

Has the technique been tested (and can it be tested)? In discussing this question, the Court noted that what distinguishes science from other fields of inquiry is reliance on empirical data to test, and potentially to falsify or refute, hypotheses derived from the theory underlying a technique's operation. At the very least, the technique should operate in a manner that allows for it to be subjected to empirical scrutiny. Better yet, it should already have been subjected to some empirical testing, and, by inference, the expert should be able to rely on the results of such testing when using the technique.

Has the technique been subjected to peer review? In listing and discussing the peer-review process as a method for gauging scientific validity, the Court recognized its unique contribution (and limitations). On the one hand, the Court noted that submission to the scrutiny of peer review increases the chances that substantive flaws in a technique will be detected. On the other, it noted that innovative techniques or analyses may be instructive even if they have yet to be subjected to peer review. Therefore, the Court indicated that publication in the peer-reviewed literature is relevant and can bolster the scientific validity of a technique, but it is not necessary or dispositive. Indeed, in the *Daubert* case itself, the newly conducted analyses at issue had yet to be subjected to peer review.

What is the known or potential rate of error associated with the technique? In guiding trial judges to consider the potential error rate associated with a challenged technique, the Supreme Court recognized (citing prior appellate court decisions on this topic) that scientifically based techniques are not (and should not be expected to be) error-free. However, information about the error rates associated with conclusions reached on the basis of the technique should be available to an expert relying on it and, when need be, communicated to the trier of fact. In this context, the Court cited a specific case having to do with the error rate of spectrographic voice identification technique (*United States v. Smith*, 869F. 2d 348, 1989), in which the United State Court of Appeals, Seventh Circuit, held that the availability of data on the reliability of this technique was sufficient to render testimony admissible on the basis of its application.

Are there standards controlling the technique's operation? In suggesting that trial judges consider the existence and maintenance of standards controlling the technique's operation, the Supreme Court cited an appellate court case (*United States v. Williams*, 583 F. 2d 1194, 1978) in which the United States Court of Appeals, Second Circuit, held that an expert should have access to and follow standard procedures for the operation of a technique. Here, the court recognized that having information about the error rate associated with use of a technique is helpful only if the technique is used in a manner consistent with how it was used when its error rate was calculated. The appellate court held that having standard operating procedures is a prerequisite for obtaining reliable information, and the Supreme Court recommended that in considering the admissibility of expert testimony trial judges determine whether standard operating procedures exist and were maintained.

Is the technique generally accepted? Although it rejected general acceptance as the sole basis for determining the admissibility of expert testimony, the Court held that it can still be considered a potential indicator of scientific validity. The Court noted that widespread acceptance can be an important factor in finding particular testimony admissible and that a known technique that has attracted only minimal support in the relevant scientific community may be viewed with skepticism.

In some states, statutory and/or case law retains the more conservative Frye test of general acceptance in the scientific community as the standard to be used by trial judges in deciding whether to admit expert testimony. In those jurisdictions, and in cases in which general acceptance is considered in the context of Daubert, the question of what constitutes general acceptance becomes critical. Survey data reflecting the frequency of use of a specific technique are frequently unavailable, and, even when they are, courts have generally been reluctant to accept “nose counts” as evidence for or against general acceptance. In one such ruling (*Ramirez v. State*, 810 So. 2d 836 - Fla: Supreme Court 2001), the Florida Supreme Court indicated:

When applying the *Frye* test, a court is not required to accept a “nose count” of experts in the field. Rather, the court may peruse disparate sources—e.g., expert testimony, scientific and legal publications, and judicial opinions—and decide for itself whether the theory in issue has been “sufficiently tested and accepted by the relevant scientific community.” In gauging acceptance the court must look to properties that traditionally inhere in scientific acceptance for the type of methodology or procedure under review—i.e., “indicia” or “hallmarks” of acceptability. (p. 844).

In applying this type of analysis to the case at hand (involving knife mark evidence), the Florida Supreme Court discussed whether the procedure in question had been tested empirically, whether it had been subjected to peer review, whether the error rate of the procedure had ever been quantified, and whether the expert in this case followed a standard procedure. In other words, the “indicia” and “hallmarks” of general acceptance applied by the Florida Supreme Court are the same as those spelled out by the US Supreme Court in its Daubert decision.

In its Daubert decision and subsequent rulings (e.g., *Kumho Tire Co. v. Carmichael*, 526 US 137, 1999), the Supreme Court indicated that an inquiry based on the Daubert factors is best viewed as a flexible one, with its ultimate goal being an assessment of the scientific validity of proffered expert testimony. It is not necessary for all of the questions laid out by the Court to be answered in the affirmative for proffered testimony to be deemed scientifically valid and therefore admissible and other questions relevant to appraisal of scientific validity may also be raised in this context. This principle is illustrated clearly in the Daubert case itself, in which the Court ruled that lower courts had erred in denying the admissibility of novel analyses of existing data, though these analyses had yet to be published in the peer-reviewed literature or to have gained general acceptance in the scientific community.

As noted earlier, a motion in limine triggering a pre-trial admissibility hearing is relatively rare, and expert testimony is more often challenged in pre-trial depositions or cross-examination during trials. In these situations, the cross-examiner is, of course, free to bring up any of the Daubert factors, as well as other questions that may reduce the weight assigned by the trier of fact to the expert's testimony. The following responses to the Daubert questions as they apply to the MMPI-2-RF would be relevant to pretrial hearings as well as to pretrial depositions and trial testimony.

Applying the Daubert Factors to the MMPI-2-RF

Has the MMPI-2-RF been tested? Recognizing the critical role played by empirical findings throughout the history of the inventory, Tellegen and Ben-Porath (2008/2011) included extensive psychometric data in the MMPI-2-RF Technical Manual. One set of findings, related to the reliability of MMPI-2-RF scores, is discussed later in the context of the known or potential error rates of the test. In this section, four other types of MMPI-2-RF data are described, focusing on booklet and normative comparability, internal structure, external correlates, and descriptive findings.

Booklet and normative comparability analyses. Empirical data reported in the MMPI-2-RF Technical Manual (Tellegen & Ben-Porath, 2008/2011) address the question of whether scores on the 51 scales are likely to be the same when derived from either the 567-item MMPI-2 booklet or the 338-item MMPI-2-RF. This is a particularly important question about the test

because all of the reliability, validity, and descriptive findings reported in the Technical Manual were derived from analyses of data collected with the MMPI-2 booklet. Results of analyses reported in Appendix C address this question. Data obtained from a sample of 490 college students tested twice were used in these analyses. A subset of 156 individuals completed the MMPI-2 twice, another 183 took the MMPI-2-RF twice, and 151 completed both the MMPI-2 and the MMPI-2-RF once. Correlations between scores on the 42 MMPI-2-RF substantive scales obtained when the MMPI-2 and MMPI-2-RF were administered to the same test takers were found to be essentially the same as those obtained when either the MMPI-2 or the MMPI-2-RF was administered twice. Comparisons of mean *T* score on the 51 scale scores (and their associated standard deviations) indicated that findings derived from the two booklets are essentially interchangeable.

Using the subset of the contemporary data (just described) that was collected with the MMPI-2-RF booklet, Appendix C of the Technical Manual also includes information needed to examine how well the test norms, collected in the mid-1980s, represent current test takers. Mean *T* scores (and their associated standard deviations) on the 51 MMPI-2-RF scales were compared for this cohort of students tested in the mid-2000s with those produced by a cohort tested at the same university when the test norms were collected in the mid-1980s. The 1980s cohort was tested with the experimental 704-item booklet used to collect the MMPI-2/MMPI-2-RF normative data. Overall, the results reported in Appendix C reflect substantial comparability of MMPI-2-RF scores generated by the two cohorts, indicating that the MMPI-2-RF *T* scores provide an adequate normative framework.

Internal structural analyses. Three sets of data pertaining to the internal structure of the MMPI-2-RF are reported in the Technical Manual. Intercorrelations of scores on the 51 scales of the inventory (a total of 1,275 correlations per sample) are reported in chapter 3 for seven samples composed of the 1,138 men and 1,138 women of the MMPI-2-RF normative sample, 410 men and 610 women tested at a community mental health center, 709 men and 473 women tested at a psychiatric inpatient unit of a community hospital, and 1,128 men assessed in a psychiatric inpatient unit of a Veterans Administration (VA) hospital. In addition, correlations between the 42 MMPI-2-RF substantive scales and the 103 substantive scales and subscales presently scored on the MMPI-2 (including the MMPI-2 Clinical, Harris-Lingoes, Content, Content Component, Supplementary, and PSY-5 Scales and Subscales) (a total of 4,326 correlations per sample) are reported for the same seven samples in Appendix E of the Technical Manual.

A final set of internal analyses are reported for data sets that were altered to generate invalid records. The validity scales of both inventories were then correlated to assess the comparability of the two sets of scales. Correlations between the measures and eight of the currently scored MMPI-2 Validity Scales (a total of 72 correlations per sample) are reported for 2,465 members of the MMPI-2 normative sample whose responses had been manipulated to simulate random responding, 2,475 members of the normative sample whose responses were manipulated to simulate fixed “true” responding, 2,475 normative sample members whose responses were manipulated to simulate fixed “false” responding, 67 psychiatric inpatients who took the test under instructions to simulate over-reporting, 86 medical patients who completed the inventory under similar over-reporting instructions, 140 personal injury litigants and disability claimants who took the MMPI-2 under standard instructions, 79 individuals who had or simulated head injuries, 82 individuals instructed to simulate under-reporting and a second sample of 192 individuals who took the test under similar under-reporting instructions.

All told, the internal structural analyses reported in the Technical Manual include 39,855 correlations between various MMPI-2-RF and MMPI-2 scale scores. These analyses, conducted with a broad range of samples representing settings in which the MMPI-2-RF is used, provide detailed information about the internal structure of the inventory.

External correlates. Correlations between scores on the MMPI-2-RF substantive scales and other information collected with large samples representing settings in which the test is intended to be used are reported in Appendix A of the Technical Manual. For samples of men and women tested in an outpatient community mental health center, correlations are reported with ratings and information provided by the intake workers who conducted the initial assessments of these individuals and with ratings provided by their therapists after they had been in treatment at the facility for approximately 1 month. For three large psychiatric inpatient samples, correlations are reported with indicators (e.g., psychosocial history, presenting symptoms and complaints, mental status on admission, initial diagnostic impressions, medication at intake and discharge) recorded by research assistants who systematically reviewed their medical charts. Correlations with variables based on chart data (including psychosocial history, legal history, mental status, and diagnostic impressions) are also reported for two large samples of men and women undergoing forensic assessments related to competency to stand trial evaluations and insanity pleas. In addition, correlations with a wide range of standard self-report measures are reported for samples of mental health and medical patients, individuals assessed at intake to a substance abuse treatment program, large samples of men and women assessed in connection with disability claims and personal injury litigation, and non-clinical samples of male and female college students.

All told, the thousands of external correlates (53,970) reported in Appendix A are based on data provided by 4,336 men and 2,337 women using 605 different criteria. These validity findings served as the primary source for identifying the empirical correlates of substantive scale scores listed in the interpretive guidelines provided in chapter 5 of the MMPI-2-RF Manual for Administration, Scoring, and Interpretation (Ben-Porath & Tellegen, 2008/2011). Correlates were included on these lists if they replicated across setting, gender, and criterion source.

Descriptive findings. Appendix D of the Technical Manual provides *Comparison Group* data, descriptive findings (mean T scores and their associated standard deviations) on the 51 MMPI-2-RF scales for a broad range of samples representing settings for which the test was intended. These data are reported for the 1,138 men and 1,138 women of the normative sample, 370 men and 582 women tested at an outpatient community mental health center, 246 men and 432 women who completed the test at an outpatient private practice, 659 men and 498 women tested at a psychiatric inpatient unit of a community hospital, 1,059 men tested in an inpatient psychiatric unit of a VA hospital, 1,151 men assessed at intake to a substance abuse treatment unit of a VA hospital, 228 men and 435 women evaluated as part of a pre-surgical evaluation for bariatric surgery, 263 men and 265 women evaluated as part of a pre-surgical screening of candidates for spine surgery or spinal cord stimulator implants, 367 men and 894 women assessed at intake to a college counseling clinic, 1,227 men and 1,989 women tested in a general college student setting, 243 men and 238 women who completed the test in connection with child custody litigation, 523 men and 480 women assessed as part of a disability claim or personal injury litigation, 463 men and 578 women tested as part of a forensic independent neuropsychological examination, 551 men and 223 women assessed in connection with pre-trial evaluations of criminal defendants, 34,933 men and 7,353 women tested at intake to a state correctional system, 988 men and 337 women tested in connection with pre-employment screening of law enforcement candidates, 4,850 men and 1,518 women assessed in connection with pre-employment screening of corrections officer candidates, and 1,289 men and 869 women tested in connection with pre-employment screening of clergy candidates.

All told the Comparison Group statistics reported in Appendix D include test protocols collected from 68,377 individuals. These setting-specific Comparison Group data are designed to facilitate consideration of a test-taker's results in the context of those of representative samples of individuals tested under similar circumstances in similar settings. For example, interpretation of an MMPI-2-RF protocol produced by a criminal defendant assessed as part of a pre-trial forensic evaluation can incorporate a comparison of that individual's results with those of a relevant cohort of individuals similarly tested. Such a comparison allows taking into account the impact of demand characteristics of the evaluation and base rates of the phenomena assessed by the test in a sample of criminal defendants.

In sum, the MMPI-2-RF has been tested extensively, and the resulting empirical findings on the psychometric properties of the inventory are reported in the Technical Manual. A similar compilation of data has never been available in a single source for any version of the inventory or for other psychological tests.

Has the MMPI-2-RF been subjected to peer review? As of August 2012, the MMPI-2-RF literature includes over 150 articles published in peer-reviewed journals. Most of these publications report empirical findings on MMPI-2-RF scale sets using data collected in various settings where the test is in use. This substantial peer-reviewed literature has accumulated in a relatively short period of time because, as just discussed, existing MMPI-2 data can be used to conduct MMPI-2-RF research and also because one set of measures, the RC Scales, was first introduced in early 2003, 5 years prior to publication of the MMPI-2-RF. An up-to-date list of MMPI-2-RF references, organized by topic, is maintained at the test publisher's website (<http://www.upress.umn.edu/test-division/MMPI-2-RF/mmpi-2-rf-references>).

Approximately a quarter of the MMPI-2-RF peer-reviewed literature focuses on the Validity Scales. Two studies provide empirical findings on measures used to detect content non-responsiveness (Dragon, Ben-Porath, & Handel, 2012; Handel, Ben-Porath, Tellegen, & Archer, 2010) and one investigation explores the utility of the under-reporting measures. Most of the investigations focus on assessment and identification of over-reporting with the MMPI-2-RF. Sixteen of these studies examine the utility of various over-reporting indicators with forensic samples (Dionysus, Denney, & Halfaker, 2011; Gervais, Ben-Porath, & Wygant, 2009; Gervais, Ben-Porath, Wygant, & Green, 2007, 2008; Gervais, Ben-Porath, Wygant, & Sellbom, 2010; Gervais, Wygant, Sellbom, & Ben-Porath, 2011; Nelson, Sweet, & Heilbronner, 2007; Rogers, Gillard, Berry, & Granacher, 2011; Sellbom, Toomey, Wygant, Kucharski, & Duncan, 2010; Smart et al., 2008; Tsushima, Geling, & Fabriga, 2011; Wygant et al., 2009, 2010, 2011; Youngjohn, Weshba, Stevenson, Sturgeon, & Thomas, 2011).

Other forensically focused peer-reviewed MMPI-2-RF publications (the publications cited here and throughout this section did not all focus on all 51 MMPI-2-RF scales, though some did) include studies with civil litigants (Gervais, Ben-Porath, & Wygant, 2009; Thomas & Youngjohn, 2010) and research related to the assessment of psychopathy (e.g., Rock, Sellbom, Ben-Porath, Salekin, 2012; Sellbom, 2011; Sellbom, Ben-Porath, & Stafford, 2007; Wygant & Sellbom, in press). Additional investigations focus on criminal defendants (e.g., Mattson, Powers, Halfaker, Akenson, & Ben-Porath, 2012;

Sellbom, Ben-Porath, Baum, Erez, & Gregory, 2008; Wygant et al., 2007), an examination of scores of individuals undergoing parental competency evaluations (Stredny, Archer, & Mason, 2006), and a descriptive study of scores generated by a sample of child custody litigants (Archer, Hagan, Mason, Handel, & Archer, 2012) and comparisons of MMPI-2-RF scores generated by parental competence evaluatees and child custody litigants (Pinsoneault & Ezzo, 2012; Resendes & Lecci, 2012).

Most of the peer-reviewed MMPI-2-RF publications did not focus specifically on forensic settings or samples, but many of these investigations can nonetheless inform forensic users of the inventory. Studies conducted in medical settings, for example, an investigation on use of the MMPI-2-RF to differentiate between patients reporting epileptic or non-epileptic seizures (Locke et al., 2010) or a study on use of the RC Scales to identify depression in patients with chronic pain (McCord & Drerup, 2011), provide findings that can be of particular relevance in civil litigation.

MMPI-2-RF research conducted in mental health settings can similarly inform use of the measure in forensic assessments when psychological dysfunction is at issue. For example, possible thought dysfunction, as indicated by various MMPI-2-RF measures, would not be expected to appear differently on the inventory in individuals undergoing forensic evaluations than it does in test-takers assessed in traditional mental health settings. Examples of such research include Arbisi, Polusny, Erbes, Thuras, and Reddy (2011), who report findings on screening for Post-Traumatic Stress Disorder and Mild Traumatic Brain Injury in a sample of National Guard troops following a 15-month deployment in Iraq; Arbisi, Sellbom, and Ben-Porath (2008) and Handel and Archer (2008), who identify psychopathology-focused empirical correlates of the RC Scales in psychiatric inpatients; Ozonoff, Garcia, Clark, and Lainhart's (2005) report of RC Scale findings in high-functioning adults with Autism Spectrum Disorders; Purdon, Purser, and Goddard's (2011) report of MMPI-2-RF over-reporting scale scores of individuals with first-episode psychosis; Sellbom, Bagby, Kushner, Quilty, and Ayearst's (2012) examination of the diagnostic construct validity of the MMPI-2-RF; and Watson, Quilty, and Bagby's (2011) study of use of the MMPI-2-RF to identify individuals with bipolar disorder.

The studies cited in this section are examples (but by no means an exhaustive list) of a broad and growing peer-reviewed literature on the MMPI-2-RF. Which and how many of these investigations will apply in a given forensic case will of course depend on the specific issues at hand. However, in general, the MMPI-2-RF has been subjected to extensive peer review.

What is the known or potential rate of error of the MMPI-2-RF? Two types of data, one more commonly found in a test manual and the other in the peer-reviewed literature, provide information relevant to addressing this question.

Reliability estimates and associated standard errors of measurement. The “Standards for Educational and Psychological Testing” (AERA, APA, NCME, 1999) call for test developers to report estimates of the reliability and standard errors of measurement (SEM) for all scales scored on a test. This information is reported in chapter 3 of the MMPI-2-RF Technical Manual (Tellegen & Ben-Porath, 2008/2011), which provides for the 51 scales of the inventory reliability estimates and their associated SEMs based on analyses of the test normative data, and several clinical samples.

In discussing this information, Tellegen and Ben-Porath (2008/2011) note the need to focus on the SEM estimates, which incorporate both the reliability estimates and consideration of scale score variance in the relevant samples. Along the same lines, it is the SEMs, not the reliability estimates, that directly address the issue of known and potential rate of error associated with a technique (i.e., scale scores). The importance of focusing on SEMs can be illustrated with some of the data reported in the Technical Manual. For example, reliability estimates for the inconsistent responding measures—Variable Response Inconsistency (VRIN-r) and True Response Inconsistency (TRIN-r)—are, as expectable, considerably lower (ranging from 0.16 to 0.52) than those obtained for the remaining validity scales. However, because scale score variances are also consistently smaller for these scales, the resulting SEMs (ranging from 7 to 9 T score points) are not appreciably higher than those reported for the remaining validity indicators. As another example, the reliability estimate for the Symptom Validity scale (FBS-r) based on Cronbach's α for normative sample men is 0.50. In the same sample, Cronbach's α for the Infrequent Responding (F-r) scale is 0.69. However, the SEM for both scales in this sample is 6 T score points, indicating comparable measurement error.

Ben-Porath (2012) notes further that higher SEM values, such as those found generally for scores on the MMPI-2-RF Validity Scales when compared with the substantive scales, indicate a need to use higher cutoffs to identify significant deviations from the norm. This is consistent with the interpretive recommendations for the MMPI-2-RF Validity Scales, which do in fact require greater deviation from the norm (than is required for identification of clinically significant elevation on the substantive scales) to raise substantial concerns about the validity of a test protocol. Exceptions to the requirement for greater deviations from the mean are the Uncommon Virtues (L-r) and Adjustment Validity (K-r) scales, for which SEM findings are comparable to those reported for the substantive scales. Tellegen and Ben-Porath (2008/2011) similarly indicate that more extreme T scores are needed to justify clinically significant inferences for the substantive scales reported to have greater SEMs.

Classification accuracy statistics. The second type of data relevant to error rate consideration involves the calculation of classification accuracy statistics for dichotomous decisions based on scale score cutoffs. For example, using mainly simulation or known group designs, many MMPI-2-RF Validity Scale studies report estimates of sensitivity and specificity and often also positive and negative predictive powers at varying base rates of invalid responding (e.g., Jones & Ingram, 2011; Rogers et al., 2011; Schroeder et al., 2012; Sellbom, Toomey, Wygant, Kucharski, & Duncan 2010; Sellbom, Wygant, & Bagby, 2012; Wygant et al., 2009, 2010).

Sensitivity is the proportion of individuals assumed to be responding invalidly who score at or above a designated cutoff on a validity scale. Specificity is the proportion of test takers assumed to be responding validly who score below the designated cutoff. Positive predictive power is the proportion of individuals who score at or above the cutoff who are assumed to have responded invalidly. And negative predictive power is the proportion of individuals who score below the cutoff who are assumed to have responded validly. Assumptions about invalid responding are based either on instructions given to research subjects (simulation designs) or indications of invalid responding based on extra-test sources (e.g., known-group classification based on performance on symptom validity tests or more elaborate, structured diagnostic criteria such as those proposed by Slick, Sherman, and Iverson, 1999, or Bianchini, Greve, and Glynn, 2005).

The complement of specificity ($1 - \text{specificity}$) is used to estimate the false positive rate (FPR) associated with a given cutoff. Investigators typically seek to keep the FPR at or below 0.10, which corresponds to specificity estimates of 0.90 or higher. Following this approach, less than 10% of individuals assumed (based on simulation or known-group research designs) to have responded validly will be mistakenly identified by the cutoff as having responded invalidly. Recall from the earlier discussion that the Daubert factor related to consideration of a technique's known or potential rate of error does not require that experts rely on error-free techniques. Rather, the Court here indicated implicitly (and in subsequent decisions explicitly) that some degree of error is to be expected and can be accepted, provided that expert witnesses have access to information that allows for consideration of the technique's known or potential rate of error in their decision-making process.

In this context, it is important to recognize that classification accuracy statistics are estimates, not absolute findings. Their accuracy depends on the soundness of the criteria used in assuming valid or invalid responding. To the extent that these designations are themselves unreliable, or not perfectly valid, the resulting statistics will underestimate classification accuracy and overestimate classification errors. Therefore, these estimates are best viewed as representing the lower-bounds of classification accuracy. This caveat applies more broadly to use of validity estimates to quantify the error associated with psychological test scores. That is, when correlations between test scores and extra-test criteria are used to estimate test score validity, it is likewise important to recognize that these statistics represent an underestimate of validity to the extent that the criteria used in calculating the correlations are themselves unreliable and otherwise fallible measures of the constructs they are intended to assess.

To summarize, reliability estimates and their associated SEMs, classification accuracy estimates, and other validation data (discussed earlier) provide abundant information about the known or potential rate of error associated with MMPI-2-RF scale scores. Proper use of this information requires an understanding of how their inherent shortcomings can result in over-estimates of error if their limitations are not considered properly when interpreting these statistical estimates.

Are there standards controlling use of the MMPI-2-RF? As reflected in its title, the “MMPI-2-RF Manual for Administration, Scoring, and Interpretation” (Ben-Porath & Tellegen, 2008/2011) provides guidelines for all aspects of using the test. Chapter 4, which outlines procedures for administration and scoring of the inventory, begins with the following admonition:

Obtaining valid information from the MMPI-2-RF requires that standard procedures be employed in the administration and scoring of the instrument. Users are encouraged to follow the procedures described in this chapter so that they can rely optimally on the normative and validation data for the MMPI-2-RF, which were collected following similar procedures. Deviation from these standard procedures jeopardizes the integrity and interpretability of the resulting test scores. The greater the deviation from standard procedures for administration and scoring, the less confidence can be placed in the validity of the test results. (p. 17)

The chapter goes on to provide detailed guidance for standard test administration and scoring of the inventory. As just noted, failure to adhere to these procedures compromises the user's ability to rely on the normative and validation data for the instrument, and hence the quality of the information obtained in an assessment. This admonition in the MMPI-2-RF manual is consistent with the Supreme Court's discussion of this particular Daubert factor (reviewed earlier), when the Court indicated that standard procedures must be available and followed to satisfy this particular criterion.

Chapter 5 of the “MMPI-2-RF Manual for Administration, Scoring, and Interpretation,” provides detailed, step-by-step guidance on how to interpret test findings. This represents a significant departure from the MMPI-2 manual (Butcher et al., 2001), which offers only a brief guide to the interpretation of test results and refers its readers to “MMPI-2 interpretive guides” for details. Owing to the diversity of approaches followed by the various authors of MMPI-2 interpretive guides, it is possible and, regrettably, not uncommon for two experts interpreting the same MMPI-2 protocol to reach rather disparate

inferences about the results. Such lack of agreement on how to interpret the same set of objective test findings can significantly limit the impact of MMPI-2-based testimony. In contrast, because the MMPI-2-RF manual includes comprehensive interpretive guidelines, two experts examining the same set of MMPI-2-RF results will produce essentially the same interpretation if they follow the standard interpretive guidelines for the test.

The improvement afforded by the availability of a comprehensive set of standard interpretive guidelines for the MMPI-2-RF is of particular relevance to forensic users of the test. Consistent with Daubert's "standards controlling the technique" criterion, it removes an element of subjectivity from the interpretation and allows experts, counsel, judges, and juries to determine whether an inference linked to the MMPI-2-RF is consistent with the standard interpretive guidelines for the inventory. It prevents situations in which dueling experts can point to an authoritative source to bolster seemingly contradictory conclusions about the same set of test results.

To summarize, standards controlling all aspects of using the MMPI-2-RF – administration, scoring, and interpretation – are spelled out in detail in the test manual (Ben-Porath & Tellegen, 2008/2011). Inclusion of detailed interpretive guidelines in this document facilitates cross-interpreter reliability and the ability of judicial fact finders to assess whether an expert has adhered to these guidelines.

Is the MMPI-2-RF generally accepted? As of the writing of this article, survey data documenting the frequency of MMPI-2-RF use have yet to be published. However, as noted previously, courts have generally shied away from relying on "nose counts" in determining general acceptance. In fact, case law in some states direct judges to consider information strongly resembling the other four Daubert Factors (empirical testing, peer review, error rate, and standard procedures) as general acceptance indicia (cf., *Ramirez v. State*, 810 So. 2d 836 - Fla: Supreme Court 2001 as discussed earlier in this article). In such jurisdictions, the MMPI-2-RF information and issues just discussed in the context of the other factors would be relevant to the determination of general acceptance.

Acceptance of the MMPI-2-RF can also be gauged by consideration of indirect indicators and by examination of published criticisms of the test.

Indications of MMPI-2-RF acceptance. Although no appellate court decision available to date has addressed directly the question of whether the MMPI-2-RF is generally accepted, courts have begun to cite MMPI-2-RF findings in their analyses. For example, in *Michigan v. Espinoza* (2011), addressing the question of whether a defendant had psychological disorders that precluded his ability to represent himself, the Michigan Court of Appeals commented "We note that the defendant was given the MMPI-2-RF as part of his evaluation [at the Center for Forensic Psychiatry], and the results did not indicate any mental illness." (p. 2). In an order denying relief for a petitioner seeking to stay his execution (*Wood v. Thaler*, 2011), a federal judge in the San Antonio Division of the US Court in the Western District of Texas cited MMPI-2-RF findings in the consideration of whether the petitioner had been feigning a mental disorder.

The authors of two leading MMPI interpretive texts (Graham, 2012; Greene, 2011) devote considerable attention to the MMPI-2-RF in the most recent editions of their books. Both authors provide detailed interpretive guidelines and case illustrations for the test. The MMPI-2-RF is also covered routinely in other psychological assessment textbooks and reference sources (e.g., Groth-Marnat, 2012; Hebben & Milberg, 2009; Kitaeff, 2011; Larrabee, 2012; Neukrug & Fawcett, 2009; Plante, 2010; Reynolds & Livingston, 2012). As of the writing of this article, the test has been the subject of over a dozen doctoral dissertation studies in the US and elsewhere. Wilde and colleagues (2010) provide a list of psychological tests recommended by a workgroup convened by the National Institutes of Health (NIH) for use in outcome research on traumatic brain injury. The MMPI-2-RF is the only personality measure included on this list. Gallo and Halgin (2011) include the test in their list of measures used in assessing law enforcement candidates.

Incorporation of the MMPI-2-RF in graduate and post-graduate continuing education instruction also provides indirect evidence of acceptance. An internet search of the terms "MMPI-2-RF syllabus" identifies a substantial and growing number of graduate courses that include training on the MMPI-2-RF. In addition, a large number of professional organizations have offered continuing education training on the instrument. A list of these organizations is provided in Table 1.

Finally, the MMPI-2-RF is being used increasingly outside the United States. The American English version is presently in use in Australia, Canada, South Africa, and the United Kingdom. Approved (by the publisher) MMPI-2-RF translations have been published in Spain (for use in Spain, South America, and Central America) and in South Korea. Adaptation projects are currently ongoing in several other languages and countries (Up-to-date information on available translations of the MMPI-2-RF can be found at the publisher's website (<http://www.upress.umn.edu/test-division/translations-permissions/permissions>)).

Table 1. Professional organizations that have provided continuing education training on the MMPI-2-RF

American Academy of Clinical Neuropsychology
American Academy of Forensic Psychology: American College of Professional Neuropsychology
American Psychological Association
American Psychology-Law Society
Arkansas Psychological Association
American Society of Metabolic and Bariatric Surgery
Association of Family and Conciliation Courts
California Psychological Association
Canadian Academy of Psychologists in Disability Assessment
Colorado Assessment Society
Dallas Psychological Association
Florida Psychological Association (Broward County)
Georgia Psychological Association (Clinical Division)
Illinois Psychological Association
Indiana Psychological Association
Iowa Psychological Association
Kansas Psychological Association
Kentucky Psychological Association
Louisiana Psychological Association
Maryland Psychological Association
Mental Health in Corrections Consortium
Michigan Psychological Association
Missouri Psychological Association
Montana Psychological Association
National Academy of Neuropsychology
Nebraska Psychological Association
New Mexico Psychological Association
New York State Psychological Association
Ohio Psychological Association
Oklahoma Psychological Association
Ontario Psychological Association
Sacramento Valley Psychological Association
San Diego Psychological Association
Society for Personality Assessment
Tennessee Psychological Association
Texas Psychological association
Utah Psychological Association

MMPI-2-RF criticisms. An expert testifying in a Daubert or Frye hearing on the admissibility of MMPI-2-RF-based testimony, in deposition, or before a judge and jury, may be confronted with, and should be prepared to address, published criticisms of the test. Ben-Porath (2012) provides a detailed discussion of early appraisals of the inventory. The main points of this discussion are reviewed here.

As noted earlier, authors of two of the major MMPI interpretive guides (Graham, 2012; Greene, 2011) provide extensive coverage of the MMPI-2-RF, which includes detailed recommendations for its use as well as their appraisals of the inventory, including some advantages and disadvantages they find in the revised test. The advantages they list include brevity, ease of interpretation, and links to the contemporary literature on personality and psychopathology. Both authors mention the loss of information from Clinical Scale code types as a potential disadvantage of the MMPI-2-RF. However, Graham (2012) notes that “one could argue that code types evolved largely as a way to deal with the heterogeneity of the Clinical Scales and are not necessary because of the homogeneity of the RC Scales and other MMPI-2-RF scales” (p. 414).

Both authors also discuss the absence of specific supplementary MMPI-2 measures as disadvantages. Graham (2012) lists the MacAndrew Alcoholism Scale–Revised (MAC-R; MacAndrew, 1965) and the Hostility Scale (Ho; Cook & Medley, 1954). He also discusses the loss of the Ego Strength Scale (Barron, 1953) with its positive focus on psychological resources. As noted earlier, the MMPI-2-RF Technical Manual (Tellegen & Ben-Porath, 2008/2011) reports correlations between MMPI-2 and MMPI-2-RF scales. Examination of these statistics indicates that MAC-R is most closely associated with the Higher-Order Behavioral/Externalizing Dysfunction (BXD) Scale of the MMPI-2-RF, and that the RC Scale Cynicism (RC3) assesses the cynical hostility component of the Ho scale. Es is a more heterogeneous measure that does not have a direct parallel in the

MMPI-2-RF. However, interpretive recommendations provided in the “MMPI-2-RF Manual for Administration, Scoring, and Interpretation” (Ben-Porath & Tellegen, 2008/2011) identify positive features associated with low scores on several MMPI-2-RF scales. This information can be used to highlight potential strengths of a test-taker.

Greene (2011) also perceives another disadvantage:

The “MMPI-2” in MMPI-2-RF is a misnomer because the only relationship to the MMPI-2 is its use of a subset of the MMPI-2 item pool, its normative group, and similar validity scales. The MMPI-2-RF should not be conceptualized as a revised or restructured form of the MMPI-2, but as a *new* self-report inventory that chose (sic) to select its items from the MMPI-2 item pool and use its normative group. (p. 22)

However, naming this instrument, made up exclusively of MMPI-2 items and standardized on the MMPI-2 norms, anything but a restructured version of the MMPI-2 would in fact be misleading. Greene (2011) goes on to write that “clinicians who use the MMPI-2-RF should realize that they have forsaken the MMPI-2 and its 70 years of clinical and research history, and they are learning a new inventory” (p. 22). Nonetheless, he provides detailed recommendations on how to use the MMPI-2-RF, which span roughly one-fourth of his book and include several case studies. Greene has also developed a commercially available computer-based interpretive report for the MMPI-2-RF. It can, therefore, reasonably be inferred that Greene does not view his expressed concerns as cause for not using the test.

A third author (Butcher, 2011) provides an exclusively negative appraisal of the MMPI-2-RF and recommends against its use. Much of Butcher’s appraisal consists of repetition of criticisms of the RC Scales without consideration of the substance of published responses to these criticisms (Tellegen et al., 2006, 2009). He also lists some new concerns, including the relatively low reliability estimates for some Specific Problems Scales. However, as discussed earlier, the reliability estimates reported in the Technical Manual need to be considered in the context of the associated measurement error statistics, which are also reported. Butcher’s (2011) claim that “the majority of the scales incorporated in the MMPI-2-RF are insufficiently validated to provide the practitioner with confidence in assessment” (p. 189) is belied by the unparalleled quantity and quality of external correlate data reported in the Technical Manual (discussed earlier). In fact, as also noted, a comparable set of validation data has not been compiled in one source and integrated into interpretive recommendations for any other version of the MMPI.

Butcher (2011) expresses concern about the loss of items related to work adjustment and treatment readiness that resulted from pruning the item pool from 567 to 338 statements. The items alluded to here are scored on two of the MMPI-2 Content Scales (Butcher, Graham, Williams, & Ben-Porath, 1990): Work Interference (WRK) and Negative Treatment Indicators (TRT). Data reported in the Technical Manual indicate that both these scales are oversaturated with demoralization variance and their distinctive features are assessed on the MMPI-2-RF with the Inefficacy (NFC) and Helplessness/Hopelessness (HLP) Scales, respectively. Moreover, as noted earlier, treatment considerations are included in the interpretive recommendations for most of the MMPI-2-RF substantive scales.

Butcher (2011) remarks that “it is likely that the interpretations and conclusions drawn from the MMPI-2-RF will differ substantially from an MMPI-2 interpretation” (p. 190) and expresses concern that this may create confusion. However, because the two MMPI versions are scored from the test-taker’s responses to the same set of items, it is unlikely that two conflicting clinical pictures will emerge. The more likely outcome is that the picture portrayed by the MMPI-2-RF may be more readily and clearly discerned. Confusion can be avoided by being clear about which version of the MMPI was used in a given assessment.

Elsewhere, Butcher (2010) is critical of use of non-gendered norms with the MMPI-2-RF, stating:

Unlike the original MMPI and MMPI-2, in which separate gender norms were provided, the MMPI-2-RF authors combined genders into one comparison sample. This situation may result in different standards being applied for men and women in assessment and prediction. Further study of this potential bias needs to be conducted. However, the MMPI-2-RF manuals do not provide the information necessary for exploring this question because raw score data by gender are not reported. (p. 14)

This criticism reflects a fundamental misunderstanding of group-specific norms. Contrary to Butcher’s assertion, gender-based norms create different standards for men and women, which can mask meaningful gender differences (cf., Reynolds & Kamphaus, 2002, 2004; Reynolds & Livingston, 2012). Non-gendered norms apply the same standard to men and women’s test scores and reflect rather than mask actual gender differences. Butcher’s (2010) assertion that the MMPI-2-RF manuals do not provide information necessary to explore this question is also incorrect. As noted earlier, means and standard deviations of scores on the 51 MMPI-2-RF scales are reported in the Technical Manual by gender for a wide range of samples, including the normative sample. Examination of these data indicates that in the normative sample mean *T* scores for the two genders are within 2 points of each other for 37 of the 51 MMPI-2-RF scales. They are within 3–4 points of each other for 9 additional scales. For the remaining 5 scales, men score 6 *T* score points higher than women on Behavioral Externalizing Dysfunction (BXD) and 8 *T* score points higher on Mechanical Physical Interests (MEC) and Disconstraint (DISC-r). Women score 6 *T* score points higher than men on Aesthetic-Literary Interests (AES) and 8 *T* score points higher on Multiple Specific

Fears (MSF). Gender-based norms would have masked these differences by setting the mean *T* score for each gender at 50. Moreover, inclusion of extensive, gender-specific descriptive data in the Technical Manual allows MMPI-2-RF users to compare a test-taker's results with samples of men and women tested in a wide range of mental health, medical, forensic, personnel screening, and non-clinical settings.

Nichols (2011) mainly repeats Butcher's (2010, 2011) criticisms, focusing mostly on his own previous (Nichols, 2006) critique of the RC scales. Detailed responses to Nichols's earlier RC Scale critiques are provided by Tellegen and colleagues (2006, 2009).

To summarize, in jurisdictions where case law indicates the need to consider Daubert-like criteria in determining general acceptance, the issues discussed in the preceding sections are also relevant to this factor. In addition, citation of MMPI-2-RF findings in legal decisions, detailed coverage of and guidance on how to use the MMPI-2-RF in the two leading MMPI textbooks, coverage of the test in recent assessment reference sources, incorporation of the instrument in graduate assessment classes, opportunities for MMPI-2-RF training offered by many national and state-level organizations, and international adaptation of the instrument all point to acceptance of the instrument. Consideration of published criticisms of the MMPI-2-RF indicates that the concerns expressed in these publications can be addressed with available empirical data. It is also worth recalling (in this context) that general acceptance does not require universal approval or preclude criticism.

Conclusion

Consideration of the MMPI-2-RF in light of the Daubert factors indicates that the five questions that can be framed by these factors can be answered affirmatively. The instrument has been subjected to extensive empirical testing. Internal correlations with MMPI-2 scales in several mental health samples, extra-test correlations with a broad range of criteria in mental health, medical, forensic, and non-clinical samples, and descriptive MMPI-2-RF findings in an even broader range of samples are reported in the MMPI-2-RF Technical Manual (Tellegen & Ben-Porath, 2008/2011). The breadth and depth of the empirical data reported in this manual are unparalleled in the documentation of other psychological tests, including previous versions of the MMPI. Availability of a broad and growing body of peer reviewed MMPI-2-RF research, all conducted within the past decade, addresses the second Daubert Factor. Reliability estimates and their associated SEMs reported in the Technical Manual, and classification accuracy statistics found in the peer-reviewed literature provide information about the known and potential rate of error associated with MMPI-2-RF scores. Standard procedures for administration, scoring, and interpretation of the inventory are detailed in the test administration manual; and adherence to these procedures facilitates cross-interpreter reliability in a manner that cannot readily be accomplished with the MMPI-2. In jurisdictions where case law identifies Daubert-like Factors as the means for gauging general acceptance, the attributes just listed are relevant to consideration of this Daubert Factor. In the absence of survey data, several indirect indicators of MMPI-2-RF acceptance can also be cited. Published criticisms of the MMPI-2-RF can be addressed with information provided in the Technical Manual and available in an extensive, modern, and actively growing peer-reviewed literature.

This article began with the observation that introduction of a new version of a psychological test brings with it challenges that can be accentuated by the adversarial nature of the legal process. In the case of the MMPI-2-RF, these challenges can be addressed by becoming familiar with the rationale for and the methods used in the revision, the information contained in the test manuals, and the growing peer-reviewed literature on the inventory.

Conflict of Interest

The author is a paid consultant to the MMPI Publisher, the University of Minnesota, and Distributor, Pearson. As co-author, he receives royalties on sales of the MMP-2-RF.

Acknowledgements

I thank Beverly Kaemmer and Auke Tellegen for their feedback on an earlier version of this paper.

References

- AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Arbisi, P. A., Polusny, M. A., Erbes, C. R., Thuras, P., & Reddy, M. K. (2011). The Minnesota Multiphasic Personality Inventory-2 Restructured Form in National Guard soldiers screening positive for posttraumatic stress disorder and mild traumatic brain injury. *Psychological Assess*, 23 (1), 203–214.

- Arbisi, P. A., Sellbom, M., & Ben-Porath, Y. S. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in psychiatric inpatients. *Journal of Personality Assessment*, 90, 122–128.
- Archer, E. M., Hagan, L. D., Mason, J., Handel, R. W., & Archer, R. P. (2012). MMPI-2-RF characteristics of custody evaluation litigants. *Assessment*, 19, 14–20.
- Barron, F. (1953). An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology*, 5, 327–333.
- Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2RF*. Minneapolis, MN: University of Minnesota Press.
- Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *MMPI-2RF: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from Cognitive Malingering Research. *The Spine Journal*, 5, 404–417.
- Butcher, J. N. (2010). Personality assessment from the 19th to the early 21st century: Past achievements and contemporary challenges. *Annual Review of Clinical Psychology*, 6, 1–20.
- Butcher, J. N. (2011). *MMPI-2: A beginner's guide (Third edition)*. Washington, DC: The American Psychological Association.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2*. 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: Manual for administration and scoring (Revised edition)*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales*. Minneapolis, MN: University of Minnesota Press.
- Cook, W. W., & Medley, D. M. (1954). Proposed hostility and Pharisaid-virtue scales for the MMPI. *Journal of Applied Psychology*, 38, 414–418.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 589 (1993).
- Dionysus, K. E., Denney, R. L., & Halfaker, D. A. (2011). Detecting negative response bias with the Fake Bad Scale, Response Bias Scale, Henry-Heilbronner Index of the Minnesota Multiphasic Personality Inventory-2. *Archives of Clinical Neuropsychology*, 26, 81–88.
- Dragon, W. R., Ben-Porath, Y. S., & Handel, R. W. (2012). Examining the impact of unscorable item responses on the validity and interpretability of MMPI-2/MMPI-2-RF Restructured Clinical (RC) scale scores. *Assessment*, 19 (1), 101–113.
- Gallo, F. J., & Halgin, R. P. (2011). A guide for establishing a practice in police preemployment postoffer psychological evaluations. *Professional Psychology: Research and Practice*, 42, 269–275.
- Gervais, R. O., Ben-Porath, Y. S., & Wygant, D. B. (2009). Empirical correlates and interpretation of the MMPI-2-RF Cognitive Complaints (COG) scale. *The Clinical Neuropsychologist*, 23, 996–1015.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment*, 14 (2), 196–208.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale and MMPI-2 Validity Scales to memory complaints. *The Clinical Neuropsychologist*, 22, 1061–1079.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2-RF over-reporting scales and RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology*, 25, 274–284.
- Gervais, R. O., Wygant, D. B., Sellbom, M., & Ben-Porath, Y. S. (2011). Associations between Symptom Validity Test failure and scores on the MMPI-2-RF validity and substantive scales. *Journal of Personality Assessment*, 93, 508–517.
- Graham, J. R. (2012). *MMPI-2: Assessing Personality and Psychopathology (5th edition)*. New York: Oxford University Press.
- Greene, R. L. (2011). *MMPI-2/MMPI-2-RF: An Interpretive Manual (Third Edition)*. New York: Allyn&Bacon.
- Groth-Marnat, G. (2012). *Integrated psychological assessment reports: Theories, guidelines, and strategies*. Hoboken, NJ: John Wiley & Sons Inc.
- Handel, R. W., & Archer, R. P. (2008). An investigation of the psychometric properties of the MMPI-2 Restructured Clinical (RC) Scales with mental health inpatients. *Journal of Personality Assessment*, 90, 239–249.
- Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010). Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment*, 22, 87–95.
- Harkness, A. R., Finn, J. A., McNulty, J. L., & Shields, S. M. (2012). The personality Psychopathology—Five (PSY-5): Recent constructive replication and assessment literature review. *Psychological Assessment*, 24, 432–443.
- Harkness, A. R., & McNulty, J. L. (1994). The Personality Psychopathology Five (PSY-5): Issues from the pages of a diagnostic manual instead of a dictionary. In S. Strack, & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 291–315). New York, NY, US: Springer Publishing Co.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis, MN: University of Minnesota Press.
- Hebben, N., & Milberg, W. (2009). *Essentials of neuropsychological assessment* (2nd ed.). New York: Wiley.
- Jackson, D. N. (1971). The dynamics of structured personality tests. *Psychological Review*, 78 (3), 229–248.
- Jones, A., & Ingram, M. V. (2011). A comparison of selected MMPI-2 and MMPI-2-RF Validity Scales in assessing effort on Cognitive Tests in a military sample. *The Clinical Neuropsychologist*, 25, 1207–1227.
- Kitaeff, J. (2011). *Handbook of Police Psychology*. New York, NY: Routledge, Taylor & Francis Group.
- Kumho Tire Co. v. Carmichael*, 526 US 137 (1999).
- Larrabee, G. J. (2012). *Forensic neuropsychology: A scientific approach*. (2nd ed.). NY: Oxford University Press.
- Locke, D. E. C., Kirlin, K. A., Thomas, M. L., Osborne, D., Hurst, D. F., Drazkowski, J. F., et al. (2010). The Minnesota multiphasic personality inventory-restructured form in the epilepsy monitoring unit. *Epilepsy and Behavior*, 17, 252–258.
- Loevinger, J. (1972). Some Limitations of Objective Personality Tests. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 45–58). Oxford, England: Academic Press.
- MacAndrew, C. (1965). The differentiation of male alcoholic out-patients from nonalcoholic psychiatric patients by means of the MMPI. *Quarterly Journal of the Studies on Alcohol*, 26, 238–246.
- Mattson, C. A., Powers, B. K., Halfaker, D., Akenson, S. T., & Ben-Porath, Y. S. (2012). Predicting drug court completion with the MMPI-2-RF. *Psychological Assessment*.

- McCord, D. M., & Drerup, L. C. (2011). Relative practical utility of the MMPI-2 RC scales versus the clinical scales in a chronic pain patient sample. *Journal of Clinical and Experimental Neuropsychology*, 33 (1), 140–146.
- Meehl, P. E. (1972). Reactions, reflections, projections. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 131–189). Oxford, England: Academic Press.
- Michigan v. Espinoza* (Mic. Ct. App. 2011).
- Nelson, N. W., Sweet, J. J., & Heilbrunner, R. L. (2007). Examination of the new MMPI-2 Response Bias Scale (Gervais): Relationship with MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology*, 29, 67–72.
- Neukrug, E. S., & Fawcett, R. C. (2009). *Essentials of Testing and Assessment (2nd edition): A Practical Guide for Counselors, Social Workers, and Psychologists*. Belmont, CA: Thompson Brooks/Cole.
- Nichols, D. S. (2006). The trials of separating bath water from baby: A review and critique of the MMPI-2 Restructured Clinical scales. *Journal of Personality Assessment*, 87, 121–138.
- Nichols, D. S. (2011). *Essentials of MMPI-2 Assessment* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Norman, W. (1972). Psychometric considerations for a revision of the MMPI. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 59–83). New York: Academic Press.
- Ozonoff, S., Garcia, N., Clark, E., & Lainhart, J. E. (2005). MMPI-2 personality profiles of high-functioning adults with Autism Spectrum Disorders. *Assessment*, 12, 86–95.
- Pinsonneault, T. B., & Ezzo, F. R. (2012). A comparison of MMPI-2-RF profiles between child maltreatment and non-maltreatment custody cases. *Journal of Forensic Psychology Practice*, 12, 227–237.
- Plante, T. G. (2010). *Contemporary clinical psychology* (3rd ed.). New York: Wiley.
- Purdon, S. E., Purser, S. M., & Goddard, K. M. (2011). MMPI-2 Restructured Form over-reporting scales in first-episode psychosis. *The Clinical Neuropsychologist*, 25, 829–842.
- Ramirez v. State*, 810 So. 2d 836 (Fla. Sup. Ct. 2001).
- Resendes, J., & Lecci, L. (2012). Comparing the MMPI-2 scale scores of parents involved in parental competency and child custody assessments. *Psychological Assessment*.
- Reynolds, C. R., & Kamphaus, R. W. (2002). *Clinical and research applications of the BASC*. NY: Guilford.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment Systems for Children, second edition*. Bloomington, MN: Pearson.
- Reynolds, C. R., & Livingston, R. A. (2012). *Mastering modern psychological testing: Theory and methods*. Boston: Pearson.
- Rock, R. C., Sellbom, M., Ben-Porath, Y. S., & Salekin, R. T. (2012). Concurrent and predictive validity of psychopathy in a batterers intervention sample. *Law and Human Behavior*.
- Rogers, R., Gillard, N. D., Berry, D. T. R., & Granacher, R. P. (2011). Effectiveness of the MMPI-2-RF validity scales for feigned mental disorders and cognitive impairment: A known-groups study. *Journal of Psychopathology and Behavioral Assessment*, 33, 355–367.
- Schroeder, R. W., Baade, L. E., Peck, C. P., VonDran, E., Brockman, C. J., & Webster, B. K. et al. (2012). Validation of the MMPI-2-RF validity scales in criterion group neuropsychological samples. *The Clinical Neuropsychologist*, 26, 129–146.
- Sellbom, M. (2011). Elaborating on the construct validity of the Levenson Self-Report Psychopathy Scale in incarcerated and non-incarcerated samples. *Law and Human Behavior*, 35, 440–451.
- Sellbom, M., Bagby, R. M., Kushner, S., Quilty, L. C., & Ayearst, L. E. (2012). Diagnostic construct validity of the MMPI-2 Restructured Form (MMPI-2-RF) scale scores. *Assessment*, 19, 176–186.
- Sellbom, M., Ben-Porath, Y. S., Baum, L. J., Erez, E., & Gregory, C. (2008). Predictive validity of the MMPI-2 Restructured Clinical (RC) scales in a batterers' intervention program. *Journal of Personality Assessment*, 90, 129–135.
- Sellbom, M., Ben-Porath, Y. S., & Stafford, K. P. (2007). A comparison of MMPI-2 measures of psychopathic deviance in a forensic setting. *Psychological Assessment*, 19, 430–436.
- Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment*, 22, 22–31.
- Sellbom, M., Wygant, D. B., & Bagby, R. M. (2012). Utility of the MMPI-2-RF in detecting non-credible somatic complaints. *Psychiatry Research*.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561.
- Smart, C. M., Nelson, N. W., Sweet, J. J., Bryant, F. B., Berry, D. T. R., Granacher, R. P., et al. (2008). Use of MMPI-2 to identify cognitive effort: A hierarchically optimal classification tree analysis. *Journal of the International Neuropsychological Society*, 14, 842–852.
- Stredny, R. V., Archer, R. P., & Mason, J. A. (2006). MMPI-2 and MCMI-III characteristics of parental competency examinees. *Journal of Personality Assessment*, 87, 113–115.
- Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *MMPI-2-RF: Technical Manual*. Minneapolis, MN: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI-2 Restructured Clinical Scales: Development, validation, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., & Sellbom, M. (2009). Construct validity of the MMPI-2 Restructured Clinical (RC) scales: Reply to Rouse, Greene, Butcher, Nichols, & Williams. *Journal of Personality Assessment*, 91, 211–221.
- Tellegen, A., Ben-Porath, Y. S., Sellbom, M., Arbisi, P. A., McNulty, J. L., & Graham, J. R. (2006). Further evidence on the validity of the MMPI-2 Restructured Clinical (RC) scales: Addressing questions raised by Rogers, Sewell, Harrison, and Jordan and Nichols. *Journal of Personality Assessment*, 87, 148–171.
- Thomas, M. L., & Youngjohn, J. R. (2010). Let's not get hysterical: Comparing the MMPI-2 validity, clinical, and RC scales in TBI litigants tested for effort. *The Clinical Neuropsychologist*, 23, 1067–1084.
- Tsushima, W. T., Geling, O., & Fabriga, L. (2011). Comparison of MMPI-2 Validity Scale scores of personal injury litigants and disability claimants. *The Clinical Neuropsychologist*, 25, 1403–1414.
- United States v. Smith*, 869 F. 2d 348 (1989).
- United States v. Williams*, 583 F. 2d 1194 (1978).

- Watson, C., Quilty, L. C., & Bagby, R. M. (2011). Differentiating bipolar disorder from major depressive disorder using the mmpi-2-rf: a receiver operating characteristics (ROC) analysis. *Journal of Psychopathology and Behavioral Assessment*, 33, 368–374.
- Wilde, E. A., Whiteneck, G. G., Bogner, J., Bushnik, T., Cifu, D. X., Dikmen, S., et al. (2010). Recommendations for the use of Common Outcome Measures in Traumatic Brain Injury Research. *Archives of Physical Medicine and Rehabilitation*, 91, 1650–1660.
- Wood v. Thaler*, 131 S.Ct. 2451 (2011).
- Wygant, D. B., Anderson, J. L., Sellbom, M., Rapiet, J. L., Algeier, L. M., & Granacher, R. P. (2011). Association of MMPI-2 Restructured Form (MMPI-2-RF) validity scales with structured malingering criteria. *Psychological Injury and Law*, 4, 13–23.
- Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 Restructured Form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group samples. *Archives of Clinical Neuropsychology*, 27, 671–680.
- Wygant, D. B., & Sellbom, M. (in press). Viewing psychopathy from the perspective of the personality psychopathology five model: Implications for DSM-5. *Journal of Personality Disorders*.
- Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. I. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology*, 22, 488–499.
- Wygant, D. B., Sellbom, M., Gervais, R. O., Ben-Porath, Y. S., Stafford, K. P., & Freeman, D. B. et al. (2010). Further validation of the MMPI-2 and MMPI-2-RF response bias scale: Findings from disability and criminal forensic settings. *Psychological Assessment*, 22 (4), 745–756.
- Youngjohn, J. R., Weshba, R., Stevenson, M., Sturgeon, J., & Thomas, M. L. (2011). Independent validation of the MMPI-2-RF somatic/cognitive and validity scales in TBI litigants tested for effort. *The Clinical Neuropsychologist*, 25, 463–476.