

## ARTICLES

# MMPI-2 and MCMI-III in Forensic Evaluations: A Survey of Psychologists

JAMES N. BOW, PhD, ABPP

*Hawthorn Center, Northville, Michigan*

JAMES R. FLENS, PsyD, ABPP

*Private practice, Brandon, Florida*

JONATHAN W. GOULD, PhD, ABPP

*Private practice, Charlotte, North Carolina*

*This study explored psychologists' procedures for administering, scoring, and interpreting the MMPI-2 and MCMI-III in forensic cases. A national sample of 137 forensic psychologists responded to an online survey that focused on their use of the MMPI-2 and MCMI-III. The survey results reaffirmed some troubling practices found in a past study with child custody evaluator, such as lack of verification of data entry, inadequate knowledge of significance cutoffs, and over-reliance on computer-generated interpretive reports. Each of these may significantly impact the accuracy of scoring and/or interpretation. Implications for forensic practice are discussed.*

**KEYWORDS** *Forensic psychology, psychological testing, MMPI-2, MCMI-III, forensic evaluations*

Numerous survey studies have focused on the use of psychological tests in forensic evaluations (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Bow, Gould, Flens, & Greenhut, 2006; Lally, 2003; Lees-Haley, 1992; Quinnell & Bow, 2001). These studies have affirmed the wide use of objective personality

---

Address correspondence to James N. Bow, Hawthorn Center, 18471 Haggerty Road, Northville, MI 48168. E-mail: bow@michigan.gov

tests. This is probably a function of the prevailing view of these tests as more reliable and valid than projective tests, with standardized methods of administration. Moreover, the scoring and interpretation of objective tests is less ambiguous and less controversial relative to projective tests. However, even the administration, scoring, and interpretation of “objective” tests may have more variability than expected, as previously found among child custody evaluators (Bow, Flens, Gould, & Greenhut, 2005). The present study explored this issue with a broad array of forensic psychologists using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2001) and Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, Millon Davis, & Grossman, 2006, 2009).

The MMPI-2 is the most frequently used personality test among forensic psychologists (Archer et al, 2006; Lees-Haley, 1992; Otto, 2002). It is widely used in a variety of forensic contexts, including personal injury (Boccaccini & Brodsky, 1999), competency to stand trial (Borum & Grisso, 1995), and child custody (Quinnell & Bow, 2001). The MMPI-2 consists of 567 true/false items and provides a variety of validity, clinical, content, supplementary, component, and research scales, assessing response style, emotional/behavioral functioning, and psychopathology. The MMPI and MMPI-2 have been widely researched, with more than 5,000 citations in the literature (Otto, 2002). Recent studies have indicated that forensic psychologists view the MMPI-2 as meeting admissibility criteria for both the *Frye* test (i.e., general acceptance; Bow et al., 2005; Lally, 2003) and the *Daubert* standard (i.e., peer-reviewed, error rate, testable, and general acceptance; Bow et al., 2005; Bow et al., 2006). It is also important to note that no successful court challenges have occurred in appellate cases when the MMPI-2 was used to assess emotional functioning or psychopathology (Otto).

Over the years, the MCMI scales have gained increased recognition and usage. The current version, MCMI-III, is commonly used in forensic evaluations (Archer et al, 2006, Bow et al., 2005; Craig, 2006; Schutte, 2001), and studies have indicated that the MCMI II/III is the second-most commonly used objective instrument in both civil and criminal evaluations (Boccaccini & Brodsky, 1999; Borum & Grisso, 1995). It consists of 175 true/false items and provides numerous scales that tap response style, personality traits/disorders, and psychopathology.

The MCMI-III is based on Millon's personality theory, and its items correspond closely with criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (4<sup>th</sup> edition; DSM-IV; American Psychiatric Association, 1994), as noted by Millon, Millon, Davis, and Grossman (2006). It is important to note that, as the MCMI-III norms are based on a clinical population; there is controversy about its use only with a “clinical” population in contrast to those who argue for its use with other forensic populations (Dyer, 2005; McCann, 2002; Schutte, 2001). The MCMI-III also uses base rate (BR) scores (range, 0–115; median, 60) to examine the probability that a person displays

the presence of a trait (BR scores 75–84) or prominence (BR  $\geq$ 85) of a syndrome or disorder. The MCMI-III is a criterion-referenced test that does not assume a normal distribution (i.e., normal curve of traits or disorders), unlike the MMPI-2, which is a norm-referenced test and does assume a normal distribution. Much controversy has surrounded the use of the MCMI-II/III, although some forensic psychologists view it as meeting *Daubert* admissibility criteria (Bow et al., 2006; Schutte).

The MMPI-2 and MCMI-III manuals provide specific instructions for administering, scoring, and interpreting each test. It is important that psychologists closely adhere to these instructions. This is particularly critical in a forensic setting wherein the findings and their interpretation may significantly impact the psycho-legal issue. Past research on child custody evaluators' practices using the MMPI-2 and MCMI-III revealed numerous areas of concern, such as unmonitored test administration, under-estimation of the test readability level, lack of verification of computer-entered data, over-reliance on computer-generated interpretive reports, use of inappropriate significance cutoffs, and dependence on supplemental/experimental scales (Bow et al., 2005). The present study will further explore these concerns among a broader group of forensic psychologists.

## METHOD

The present survey was an extension of a previous study by the authors that focused on the administration, scoring, and interpretation of the MMPI-2 and MCMI-III among child custody evaluators (Bow et al., 2005). The authors decided to expand the study to include all types of forensic psychologists. Additional survey items were included to assess areas not typically analyzed in child custody evaluations, such as malingering.

The researchers utilized the SurveyMonkey Internet survey program. At the beginning of the survey, the purpose of the study was reviewed, and informed consent information was provided. It was also noted that the appropriate institutional review board had approved the study.

E-mail addresses for potential participants were obtained from public access referral sites for forensic psychologists, specifically the American College of Forensic Examiners; the American College of Forensic Psychology; and the American Board of Forensic Psychology. Each potential participant received an e-mail describing the study, including a link to the survey. Potential participants who did not respond within 2 and 4 weeks were contacted with a follow-up e-mail, unless they had indicated that they declined to participate in the study.

The e-mail and survey link were received by 653 potential respondents. A total of 137 respondents completed the survey, whereas 43 declined to

participate. This provided a return rate of 21%. No respondents from the original child custody study were included in this sample.

Respondents represented 32 states. Their average age was 56 (*standard deviation* [*SD*] 7.86), with a mean of 17.86 years (*S.D.* 8.17) of forensic experience. The vast majority of respondents were male (73%) and almost all were Caucasian (92%). Ninety-eight percent of the respondents held a doctoral degree, and the overwhelming majority (82%) worked in private practice. As a group, respondents reported that they devoted an average of 57% of their practice to forensic work. Respondents reported that they conducted more than one type of forensic evaluation, such as competency (71%); criminal responsibility (61%); child custody (53%); personal injury (50%); disability (46%); juvenile disposition (39%); parental termination (34%); death penalty (25%); corrections (25%); civil commitment (23%); workplace violence (23%); and sexual harassment (19%).

## RESULTS

Each test will be described separately; however, a comparison of major findings between the tests is provided in Table 1.

### MMPI-2

One hundred nineteen of the respondents used the MMPI-2 in forensic evaluations, which comprised 87% of the total sample. Of those respondents using the test, the median range of number of MMPI/MMPI-2s administered was 201 to 250. Respondents almost always (95%) used the full test (567 items) rather than an abbreviated version, and standard instructions were used by 84% of respondents. Eighty-eight percent of the respondents reported administering the test in an office setting; only 12% indicated giving it in a lobby or waiting room. None allowed the examinee to take the test home. Eighty-six percent of respondents used the paper and pencil administration, and 14% of respondents had examinees input the data into a computer.

Respondents were queried about the required reading level for the MMPI-2. Findings revealed a bi-modal distribution, with an equal number (35%) identifying the sixth grade and eighth grade. Nine percent identified a reading level below the sixth grade, and 10% responded that the test required a reading level above eighth grade. The remaining respondents (12%) indicated that the required reading level was seventh grade. To determine an examinee's reading level, respondents reported using one or more of the following methods: 72% relied on educational history, 60% used an informal reading assessment (i.e., reading a few items), 58% used a formal assessment (i.e., reading test), and 2% used an IQ test.

**TABLE 1** Comparison of Administration, Scoring, and Interpretation of MMPI-2 and MCMI-III: Forensic Psychologists Versus Child Custody Evaluators

Specific factors	Forensic psychologists		Child custody evaluators*	
	MMPI-2	MCMI-III	MMPI-2	MCMI-III
% Using test	87	55	98	63
% Using standardized instructions	84	91	79	93
Location of administration				
% Office	88	84	75	72
% Lobby/waiting room	12	16	25	28
% Take home	0	0	0	0
% Using paper and pencil version	86	81	90	85
Identified reading level	6 <sup>th</sup> (35%) & 8 <sup>th</sup> (35%)	9 <sup>th</sup> (40%)	8 <sup>th</sup> (32%)	6 <sup>th</sup> (38%)
Scoring paper and paper version				
% Data entry via keypad	45	46	49	56
% Mail-in service	22	27	32	32
% Hand-scored	18	13	10	8
% Scanner	15	14	9	4
Verifying keypad data entry	63	64	65	85
Report format				
% Scores only	39	21	53	38
% Interpretive report	61	79	47	62
Using interpretive reports				
% Who wrote their own descriptions based on report	57	56	59	64
% Who slightly modified statements from report	23	23	22	14
% Who copied statements verbatim from report	8	8	9	8
% Who used scores only to write own descriptions	12	13	11	14

\*Bow, J. N., Flens, J. R., Gould, J. W., & Greenhut, D. (2005). An analysis of administration, scoring, and interpretation of the MMPI-2 and MCMI-III in child custody evaluations. *Journal of Child Custody*, 2, 1-22.

For scoring the paper-and-pencil MMPI-2 form, 45% of respondents entered the data into the computer using the keypad, 22% used mail-in service, 18% hand-scored the test, and 15% used a scanner. Of those who personally input the data, 63% used the verification mode (i.e., reentering the data twice) to ensure that the data were correctly entered. Pearson Assessment (formerly NCS) was almost universally used by all respondents (88%), with the Caldwell Report being used by 8%. Three percent used their own scoring program, and another 2% used other systems.

Concerning the report format, 39% of respondents used only the computer to calculate scores whereas 61% relied on computer-generated interpretive reports. Those who used interpretive reports indicated that they used the following types in their practice (more than one type might have been indicated): Adult Clinical System by James Butcher/Pearson Assessment (40%);

Reports for Forensic Settings by James Butcher/Pearson Assessment (31%); Personnel System by James Butcher/Pearson Assessment (16%); Criminal Justice and Correction Report by Edwin Megargee/Pearson Assessment (8%); Alex Caldwell Report (10%); Alex Caldwell Child Custody Report (4%); Alex Caldwell Personnel Report (3%); and Roger Greene PAR Report (7%).

When reporting findings from interpretive reports, 57% of respondents indicated that they wrote their own descriptions based on the report, 8% indicated that they copied sentences or paragraphs verbatim, and 23% made slight modifications before using the sentences and paragraphs from the report. The remaining respondents (12%) reported that they used only the scores and wrote their own descriptions.

Respondents who did not use interpretive reports were queried about their reason(s); a list of possible reasons was provided, and respondents usually marked more than one. Eighty-nine percent said they felt sufficiently knowledgeable to write their own summary of the findings, 70% expressed concern about how the statements were generated, 39% expressed concerns about attorneys obtaining the interpretive report via discovery, 29% indicated that it was too costly, and 18% said the reports lacked context-specific norms.

Table 2 shows the respondents' ratings of the utility of different MMPI-2 validity scales for assessing defensiveness and malingering. Concerning defensiveness, the K-correction and Lie scales were rated by respondents as most useful, both receiving mean ratings in the high fives on a Likert scale from 1 (useless) to 7 (extremely useful). The next-most useful scale was

**TABLE 2** Usefulness of MMPI-2 Validity Scale for Assessing Defensiveness and Malingering

Scale	Defensiveness		Malingering	
	Mean	SD	Mean	SD
True response inconsistency	4.19	1.80	4.29	1.87
Variable response inconsistency	4.34	1.85	4.46	1.97
Infrequency	4.95	1.62	5.56	1.46
Back infrequency	4.43	1.69	5.15	1.56
Infrequency psychopathology	4.40	1.75	5.32	1.69
Lie	5.76	1.40	4.98	1.65
K-correction	5.89	1.26	4.74	1.71
Superlative	4.55	1.66	3.84	1.76
Can't say	4.04	1.92	3.69	1.81
Impression management	3.94	1.76	—	—
Self-deception management	3.78	1.68	—	—
Subtle obvious	3.74	1.73	4.04	1.85
F-K	4.29	1.63	4.17	1.71
Fake bad	—	—	5.01	1.68
Dissimulation	—	—	4.70	1.65

SD = standard deviation.

Note: Each scale was rated on a Likert scale (1 = Useless to 7 = Extremely Useful).

Infrequency (F), which received an average rating of 4.95. For malingering, the Infrequency (F, 5.56) scale was rated most useful, followed by Infrequency Psychopathology (5.32), Back Infrequency (5.15), Fake Bad (5.01), and Lie (4.98).

Respondents rated the usefulness of clinical scales using the same Likert scale (1 = useless to 7 = extremely useful). As shown in Table 3, the Psychopathic-Deviate scale received the highest mean rating, followed by Depression, Paranoid, and Mania, all with ratings of 5.50 or higher. The scale viewed as least useful was Masculine-Femininity, which received a rating of 3.34.

Respondents were also queried about their usage of other MMPI-2 scales and their utility (Table 4). The Content and Supplementary scales were used by at least 80% of respondents, and 70 % of respondents used the Harris-Lingoes subscales. Interestingly, the Restructured Clinical and PSY-5 scales received the lowest usage rates, with fewer than half of the respondents reporting using them. Concerning utility, the Content scales received the highest rating, followed by the Harris-Lingoes and Supplementary scales, with scores in the mid- to low 5 range.

**TABLE 3** Usefulness of MMPI-2 Clinical Scales in Forensic Evaluations

Scale	Mean	SD
Hypochondriasis	4.78	1.46
Depression	5.61	1.31
Hysteria	4.85	1.41
Psychopathic-deviate	5.75	1.24
Masculinity-femininity	3.24	1.57
Paranoid	5.59	1.12
Psychasthenia	5.14	1.21
Schizophrenia	5.45	1.40
Hypomania	5.59	1.29
Social introversion	4.42	1.40

*Note:* Each scale was rated on a Likert scale (1 = Useless to 7 = Extremely Useful).

**TABLE 4** Utilization and Usefulness of Other MMPI-2 Scales

Scale	Utilization		Usefulness	
	Yes (%)	No (%)	Mean	SD
Content	88	12	5.44	1.37
Content component	59	41	4.93	1.73
Supplementary	81	19	5.37	1.32
Harris-Lingoes subscales	70	30	5.39	1.41
Restructured clinical	36	64	4.59	1.86
PSY-5	42	59	4.38	1.87

SD = standard deviation.

*Note:* Each scale was rated on a Likert scale (1 = Useless to 7 = Extremely Useful).

Twenty-five percent of respondents reported hand-scoring other supplementary scales. Sixteen percent of respondents reported plotting the MMPI-2 profile on the MMPI scoring sheet, and 25% reported using non-gender T-scores. A majority of respondents reported using non-K-corrected scores (62%) and context-specific norms (60%).

Regarding the significance cutoff for interpreting the MMPI-2, 63% of respondents indicated that they used a T-score of 65, whereas 11% and 19% used a T-score cutoff of 60 and 70, respectively. Five percent of respondents indicated that they used a cutoff of 75 or higher, and 2% used a cutoff of 50.

The time frame for re-administering the MMPI-2 was explored as well. A near-majority of respondents (49%) indicated that they wait 6 to 12 months, and 33% revealed that they re-administer the test less than 6 months after the original evaluation. Fifteen percent of respondents reported waiting 13 to 19 months, 2% waited 19 to 24 months, and 1% waited more than 24 months.

Regarding admissibility issues, 99% of respondents viewed the MMPI-2 as meeting the *Frye* test, whereas 94% saw it as meeting the *Daubert* standard. Seventeen percent of respondents reported having been challenged about the admissibility of the MMPI-2.

### MCMII-III

Fifty-five percent of respondents reported using the MCMII-III in forensic evaluations. As a group, these respondents reported an average of 14.76 years (*SD* 6.20) of experience using this test. The median range of number of MCMIIs given was 101 to 150. Participants' responses as to the required reading level for this test varied greatly. The most common response was ninth grade (40%), followed by seventh grade (25%), and eighth grade (15%). Ten percent of respondents indicated that the reading level was sixth grade or lower, and another 10% thought it was tenth grade or higher. In terms of assessing reading level, respondents indicated that they used one or more of the following methods: educational history (76%), informal reading assessment, such as reading a few items (68%), formal reading assessment or reading test (54%), and IQ measure (3%).

Eighty-four percent of respondents reported administering the MCMII-III in an office setting, whereas 16% used the lobby or waiting room. None of the respondents acknowledged allowing the examinee to take the test at home. Ninety-one percent of respondents reported using the standard instructions, with 9% reporting that they modified them.

The paper-and-pencil version of the test was used by 81% of respondents, with the remaining respondents using a computer for administration. Of those using the paper-and-pencil version, 46% of respondents scored the test by entering data into the computer-scoring program via keypad, 27% used the mail-in scoring service, 14% used a scanner, and 13% hand-scored

the test. Of those entering data via keypad, 64% reported using the scoring verification feature (i.e., reentering data twice to eliminate errors).

Respondents almost universally (98%) used Pearson Assessment as the scoring/interpretive service. Regarding the type of printouts utilized, 79% of respondents used the interpretive report, with the remaining number of respondents using the profile report (scores only). Of those using interpretive reports, 96% used the Pearson Assessment Report, 2% used Robert Craig's PAR Report, and another 2% used an unidentified report. When using the interpretive report, 56% of respondents indicated that they write their own descriptions based on the information in the report. However, 23% of respondents reported inserting slightly modified sentences and paragraphs from the report into their own forensic reports, and 8% of respondents indicated that they insert entire sections verbatim. Thirteen percent of these respondents claimed that they use only the scores from the interpretive report and write their own description, which is surprising considering the greater cost of ordering the interpretive report.

Those respondents who did not use the interpretive report were queried as to their reasons; a list of possible reasons was provided, and respondents usually indicated more than one. Eighty-five percent indicated that they were sufficiently knowledgeable to interpret the data without the interpretive report, 69% indicated that the interpretive reports exaggerated psychopathology, 69% expressed concerns that attorneys might discover and gain access to the interpretive report, 62% expressed concerns about how the interpretive statements were generated, 54% complained about the cost of interpretive reports, 7% complained about gender-biased concerns, and 7% complained about vitriolic language.

Respondents were queried about the cutoff score they used for interpreting BR scores on the MCMI-III. Only 51% of respondents correctly identified the significance base-rate cutoff of 75. Interestingly, 42% thought the BR cutoff was 70 or lower. Seven percent of respondents indicated that the BR cutoff score was 50, 13% reported 60, 15% indicated 65, and 7% revealed 70. Regarding the prominence BR cutoff, only 29% of respondents correctly endorsed 85. Additionally, BRs selected ranged from 40 through 90, and 69% of respondents gave an answer of 80 or less.

Thirty-five percent of respondents thought the MCMI-III suffered from gender bias. Those respondents identified the following three scales as most affected: Histrionic, Narcissistic, and Dependent.

Respondents rated the usefulness of the MCMI-III modifying indices and personality domains on a Likert scale (1 = useless to 7 = extremely useful). The following mean scores were obtained: Modifying Indices, 4.98; Clinical Patterns, 5.48; Severe Personality Patterns, 5.51; Clinical Syndromes, 5.41; and Severe Clinical Syndromes, 5.35. All of these scores fell within the moderately useful range.

Regarding re-administering the MCMI-III, 58% of respondents indicated that they would wait 6 to 12 months, whereas 27% of respondents reported that they would do so after a period of less than 6 months. Twelve percent of respondents reported that they would wait 12 to 18 months, and the remaining respondents would wait more than 24 months.

Seventy-six percent of all respondents believed that the MCMI-III met the *Frye* test, and 69% of all respondents indicated that it met *Daubert* criteria. Twenty-one percent of respondents using the MCMI-III reported experiencing an admissibility challenge.

## DISCUSSION

The present study explored the administration, scoring, and interpretation of the MMPI-2 and MCMI-III with a wide array of forensic psychologists, extending a previous study by the authors that focused exclusively on child custody evaluators. Findings from the 137 respondents indicated that a high percentage of them used the MMPI-2 (87%) and MCMI-III (55%) in forensic work. These findings affirm previous survey research (Archer et al., 2006). The MMPI-2 remains the most widely used and popular test on the market. This is probably a function of the vast body of research on the MMPI/MMPI-2 and its wide acceptance (Otto, 2002; Pope, Butcher, & Seelen, 2006). In the present study, 99% and 94%, respectively, of respondents rated the MMPI-2 as meeting the *Frye* test and *Daubert* standard. However, 17% of respondents reported having experienced an admissibility challenge involving the MMPI-2.

The MCMI-III continues to be used by more than half of forensic psychologists, although the test has faced much criticism (Hsu, 2002; Rogers, Salekin, & Sewell, 1999, 2000). These criticisms have focused on the instrument's scientific validity and error rates. Rogers and colleagues have also argued that the MCMI-III does not meet the *Daubert* standard. Surprisingly, in the present study, 69% of respondents indicated that the MCMI-III did meet this standard; additionally, 76% of respondents viewed it as passing the *Frye* test. Moreover, the admissibility challenge rate was similar to the MMPI-2.

Comparable findings were found among the tests in the present study, as highlighted in Table 1. Moreover, many of the findings in the present study (e.g., forensic psychologists) were similar to the findings in the authors' previous study with child custody evaluators (Bow et al., 2005). The paper-and-pencil version was the preferred method of administration. The tests were almost always given in an office setting; only a small percentage of respondents allowed the examinees to take the test in the lobby, and none allowed examinees to take the test home. The latter would be absolutely taboo in a forensic setting, and even the former is strongly discouraged in light of the lack of adequate monitoring (Butcher et al., 2001). Respondents

almost always used the standard instructions for each test, which is critical to maintain standardization.

The variation of reading levels identified as necessary for the different tests reflected the ongoing debate in the field. The recommended reading level for the MMPI-2 has ranged from a fifth- to eighth-grade level, although the present authors have previously argued for the necessity of an eighth-grade reading level (Bow, Flens, & Gould, 2006). For the MCMI-III, respondents in the present study indicated a much higher reading level, with 40% rating it as ninth grade and 65% rating it as an eighth-grade level or higher. The items on the MCMI-III are particularly difficult in light of its vocabulary and its inclusion of double-negatives. It is important to consider these factors when administering the MCMI-III, and the authors would recommend at least an eighth-grade reading level as cited in the manual (Millon et al., 2006).

Concerning scoring, the overwhelming majority of respondents used computer scoring or interpretive services, almost all of which were approved services, with Pearson Assessment being the most popular. However, of those who entered the data via the computer keypad, only about 60% verified their entries. This result resembles findings from the authors' previous study with child custody evaluators (Bow et al., 2005; see Table 1). In a testing setting that requires a high degree of scrutiny, it is imperative that forensic examiners verify all data entry. It is also troubling that 18% and 13%, respectively, of respondents in this study hand-scored the MMPI-2 and MCMI-III. These figures are even higher than the study with child custody evaluators (Bow et al.). The probability of errors increases significantly with hand-scoring, rendering it an unwise practice in forensic cases (Bow et al.).

One of the most disturbing findings of this study was the high usage rate of computer-generated interpretive reports, with 61% and 79%, respectively, of respondents using such reports with the MMPI-2 and MCMI-III. As shown in Table 1, these findings are about 15 percentage points higher than the child custody study. The high usage rate of the MCMI-III interpretive reports is particularly alarming because Pearson Assessment (formerly NCS) uses a BR cutoff of 60 for descriptive statements in such reports, whereas the clinical significance BR cutoff is 75 (Millon et al., 2006, 2009). Moreover, a BR of 60 is the median score, meaning that half of the individuals in the normative sample scored at or above that score. Consequently, the MCMI-III interpretive reports are likely to over-diagnose psychopathology. Even among strong advocates of the MCMI-III, Schutte (2001) recommends that forensic psychologists exercise caution when using the MCMI-III interpretive report, whereas McCann (2002) stated that such reports should be avoided.

Flens (2005) identifies a variety of concerns regarding computer-generated test interpretations, including (1) Is the program actuarial or automated?

(2) Does the program consider response style when offering statements? (3) What level of significance is used for cutoff scores? (4) Are different statements used depending on the degree of elevation? (5) Does the program consider the profile configuration or combination of elevated scales or are statements based on single-scale elevations? and (6) Does the program use context-specific normative data? Flens stalwartly argues that such interpretive reports should not be used because psychologists do not know how each interpretive statement was derived.

Interestingly, this problem has been rectified with the new Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2 RF; Ben-Porath & Tellegen, 2008). The computer-generated interpretive report for the MMPI-2 RF provides citations for each interpretive statement, thereby outlining the rationale. This change in format may have been the publisher's response to these ongoing criticisms.

The most surprising finding was respondents' lack of knowledge about significance levels for the MMPI-2 and MCMI-III. For the MMPI-2, only 63% of respondents reported using the recommended significance T-score cutoff of 65. The remaining respondents used T-score cutoffs ranging from 50 to 75, with 19% indicating that they used a T-score cutoff of 70. Interestingly, the latter was the significance cutoff for the original MMPI. For the MCMI-III, only 51% of respondents knew that the significance BR cutoff was 75. Thirty-five percent of respondents indicated that it was 65 or less. Furthermore, fewer than one-third of respondents knew that the prominence BR cutoff was 85. The previous study with child custody evaluators found slightly higher figures, with 60% of respondents identifying the correct BR cutoff as 75 and 53% identifying the prominence cutoff as 85 (Bow et al., 2005). These findings raise serious questions about the ability of many psychologists to interpret these tests and might explain their reliance on computer-generated interpretive reports. In the final analysis, psychologists have an ethical obligation to thoroughly understand the tests that they score and interpret.

On the MMPI-2, the Validity and Clinical scales continue to be the mainstays of the test. The Restructured Clinical and PSY 5 scales were utilized by less than half of the respondents and were rated as only moderately useful. In many ways, these findings were expected, considering the relative newness of these scales and lack of relevant empirical research. The Content and Supplementary scales and Harris-Lingoes subscales were used by significantly more respondents (70% or more) and have a more solid research basis. To assess defensiveness, respondents identified the Lie and K-Correction as the most useful Validity scales. Both of these scales were original validity scales developed to assess under-reporting and have much research support (Graham, 2006). For malingering, the Infrequency (F), Infrequency-Psychopathology, Fake Bad, Back Infrequency, and Lie were rated most useful. Only the latter was an original validity scale; the other

scales were later developed to assess over-reporting or malingering. A meta-analysis of MMPI-2 feigning studies indicated that the Infrequency and Infrequency-Psychopathology were the most effective scales, but the latter scale was most preferred (Rogers, Sewell, Martin, & Vitacco, 2003).

Among the MMPI-2 Clinical scales, the Psychopathic-Deviate, Depression, Paranoid, and Hypomania were rated as most useful. This finding was somewhat expected considering the types of psychopathology assessed in forensic evaluations. However, the omission of the Schizophrenia scale among this group was puzzling. This might be a function of the weaker diagnostic properties of the scale (Freidman, Lewak, Nichols, & Webb, 2001).

Interestingly, the use of MMPI-2 context-specific norms was endorsed by 60% of respondents. In addition, of those using computer-generated interpretive reports, 41% of respondents reported using the forensic version. This is an important trend in the field of forensic psychology. Context-specific norms have been available for years in the child custody field, but only recently have these norms been expanded to other areas of forensic psychology. It is important to utilize such context-specific norms when they are available, and it is hoped that future research will focus on developing such norms for different forensic populations.

For the MCMI-III, respondents rated the modifying indices and personality and syndrome scales as moderately useful, ratings that resembled those of the MMPI-2 main clinical scales. Interestingly, although the MCMI-III is often considered a better measure of Axis II disorders (i.e., personality disorders) than of Axis I disorders/syndromes (i.e., major psychopathology; (Millon et al., 2006, 2009), respondents rated the test as similarly useful in both areas.

Thirty-five percent of respondents indicated concern about gender bias with the MCMI-III. This issue may have been resolved with the recent publication of non-gender norms for the MCMI-III (Millon et al., 2009), but further research is needed.

Concerning retesting, the vast majority of respondents reported that they would wait at least 6 months before re-administering the MMPI-2 or MCMI-III. This is an interesting finding as many litigants are tested numerous times throughout the course of their legal case. Furthermore, this provides some guidance for forensic psychologists regarding the retest issue.

One limitation of this study is that the findings reflect a sample composed of an experienced group of forensic psychologists, the overwhelming majority of whom worked in a private practice setting. Therefore, the findings may not apply to forensic psychologists who are less experienced or work in other settings. Also, the present study focused only on the MMPI-2 and MCMI-III and may not reflect forensic psychologists' administration, scoring, and interpretive procedures with other tests. Last, although the response rate was relative low (21%), research has indicated that Internet Web-based surveys have a lower response rate than traditional mail surveys (Cook, Heath, & Thompson, 2000).

## SUMMARY AND IMPLICATIONS FOR PRACTICE

In general, the findings of the present study with forensic psychologists were comparable to a previous study that focused exclusively on child custody evaluators. This is interesting because forensic psychologists are often viewed as having greater expertise in forensic methods and procedures. The present study again highlighted ongoing concerns about the accuracy of scoring, over-reliance on computer-generated reports, and lack of knowledge about the significance cut-off for each test. These findings have serious implications for forensic practice.

In light of the potential impact of test findings on the psycho-legal issue, it is critical that examiners ensure that their work product reflects high-quality, ethically based, and empirically supported procedures. The following implications for practice flow from the authors' present and past research findings (Bow et al., 2005, 2006) as a means of attaining this goal:

- Examiners should obtain prerequisite training before using any psychological test.
- Examiners should thoroughly read and understand the manual for each test.
- Examiners should be aware of the pertinent research regarding each test, including the review in *Buros Mental Measurements*.
- Examiners should use only officially published tests and protocols.
- Examiners should make sure the examinee's reading level is commensurate with the reading level of the test, using a conservative estimate for the latter. For the MMPI-2 and MCMI-III, it is strongly recommended that an eighth-grade reading level be ascertained.
- Examiners should use the standard instructions for each test.
- Examiners should administer all objective tests in a quiet office setting, with frequent monitoring.
- Examiners should score objective tests via computer using approved programs.
- Examiners should verify data entry when inputting data via computer keypad.
- Examiners should use extended scoring programs rather than computer-generated interpretive reports.
- Examiners should follow their own empirically based interpretive decision rules, which they can clearly describe in the report and during testimony.
- Examiners should be sure to address response style issues, particularly their possible impact on the findings.
- Examiners should use the significance cutoff score identified in the manual. For the MMPI-2, the T-score is 65, whereas for the MCMI-III the BR score is 75. When examiners deviate from these cutoff scores, they must

provide a clear explanation for this and cite the relevant research to support their decision.

- Examiners need to be aware of context-specific norms and of K-correction issues in defensive profiles.
- It is important for examiners to remember that descriptors about specific elevations or high point pairs are reflective of group findings, not of a particular individual. Therefore, results should be explained as, “Individuals with a similar elevation (or pattern) may exhibit the following .....
- It is important for examiners to remember that specific elevations or high-point pairs are reflective of dimensional rather than categorical descriptors.
- Examiners need to remain within the limits of the data and to avoid over-interpreting.
- It is critical to remember that test findings are used to generate and confirm hypotheses and should never be used in isolation. Testing is only one component of a forensic evaluation.

Through following these guidelines, it is hoped that forensic psychologists will be better able to serve litigants and the court with regard to psychological testing.

## REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed.). Washington, DC: Author.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological tests use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*, 85–95.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2 RF: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic test use by forensic psychologists emotional injury cases. *Professional Psychology: Research and Practice*, *30*, 253–259.
- Borum, R., & Grisso, T. (1995). Psychological tests use in criminal forensic evaluations. *Professional Psychology: Research and Practice*, *26*, 465–473.
- Bow, J. N., Flens, J. R., Gould, J. W., & Greenhut, D. (2005). An analysis of administration, scoring, and interpretation of the MMPI-2 and MCMI-II/III in child custody evaluations. *Journal of Child Custody*, *2*(4), 1–22.
- Bow, J. N., Gould, J. W., Flens, J. R., & Greenhut, D. (2006). Testing in child custody evaluations- Selection, usage, and *Daubert* admissibility: A survey of psychologists. *Journal of Forensic Psychology Practice*, *6*, 17–38.
- Bow, J. N., Flens, J., & Gould, J. W. (2006). Commentary: MMPI-2 readability. *Journal of Child Custody*, *3*, 71–75.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2 Manual for*

- Administration, Scoring, and Interpretation* (rev. ed.). Minneapolis: University of Minnesota Press.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-based or Internet-based surveys. *Educational and Psychological Measurement, 60*, 821–836.
- Craig, R. J. (2006). The Millon Clinical Multiaxial Inventory-III. In R. P. Archer (Ed.), *Forensic uses of clinical assessment instruments* (pp. 121–145). Mahwah, NJ: Lawrence Erlbaum.
- Dyer, F. J. (2005). Forensic applications of the MCMI-III in light of current controversies. In R. J. Craig (Ed.), *New Directions in Interpreting the Millon Clinical Multiaxial Inventory-III*. (pp. 201–223). New York: John Wiley & Sons.
- Flens, J. R. (2005). The responsible use of psychological testing in child custody evaluations: Selection of tests. *Journal of Child Custody, 2*, 3–29.
- Freidman, A. F., Lewak, R., Nichols, D. S., & Webb, J. (2001). *Psychological assessment with the MMPI-2*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Graham, J. R. (2006). *MMPI-2: Assessing Personality and Psychopathology* (4<sup>th</sup> ed.). New York: Oxford Press.
- Hsu, L. M. (2002). Diagnostic validity statistics and the MCMI-III. *Psychological Assessment, 14*, 410–422.
- Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice, 34*, 491–498.
- Lees-Haley, P. (1992). Psychodiagnostic test usage by forensic psychologists. *American Journal of Forensic Psychology, 10*, 25–30.
- McCann, J. T. (2002). Guidelines for forensic application of the MCMI-III. *Journal of Forensic Psychology Practice, 2*, 55–69.
- Millon, T., Millon, C., Davis, R., & Grossman, S. (2006). *MCMI-III Manual* (3<sup>rd</sup> ed.). Minneapolis: NCS Pearson.
- Millon, T., Millon, C., Davis, R., & Grossman, S. (2009). *MCMI-III Manual* (4<sup>th</sup> ed.). Minneapolis: NCS Pearson.
- Otto, R. (2002). Use of the MMPI-2 in forensic settings. *Journal of Forensic Psychology Practice, 2*, 71–91.
- Pope, K. S., Butcher, J. N., & Seelen, J. (2006). *The MMPI, MMPI-2, and MMPI-A in Court: A Practical Guide for Expert Witnesses and Attorneys* (3<sup>rd</sup> ed.). Washington, DC: American Psychological Association.
- Quinnell, F. A., & Bow, J. N. (2001). Psychological tests used in child custody evaluations. *Behavioral Science and the Law, 19*, 491–501.
- Rogers, R., Salekin R. T., & Sewell, K. W. (1999). Validation of the Millon Clinical Multiaxial Inventory for Axis II disorders: Does it meet the *Daubert* standard? *Law and Human Behavior, 23*, 425–443.
- Rogers, R., Salekin R. T., & Sewell, K. W. (2000). The MCMI-III and the *Daubert* standard: Separating rhetoric from reality. *Law and Human Behavior, 24*, 501–506.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 160–177.
- Schutte, J. W. (2001). Using the MCMI-III in forensic evaluations. *American Journal of Forensic Psychology, 19*, 5–20.