An Analysis of Administration, Scoring, and Interpretation of the MMPI-2 and MCMI-II/III in Child Custody Evaluations

James N. Bow
James R. (Jay) Flens
Jonathan W. Gould
Dana Greenhut

James N. Bow, PhD, is Director of Psychology at Hawthorn Center and adjunct Assistant Professor at Wayne State University, School of Medicine, Department of Psychiatry and Behavioral Neurosciences. He also has a private forensic practice at University Psychiatric Center-Livonia. Dr. Bow has published a variety of articles on child custody practice. He may be contacted through his forensic Website: <www.drjamesbow.com>.

James R. Flens, PsyD, is a clinical and forensic psychologist in private practice in Brandon, Florida, and a principal with Child Custody Consultants, a national consulting consortium. His practice is devoted to child custody related issues, including evaluations, expert testimony and consultation with attorneys. Dr. Flens is also involved with many research projects involving psychological testing in the child custody context, and serves on the editorial board of the *Journal of Child Custody*.

Jonathan W. Gould, PhD, is a practicing forensic psychologist specializing in child custody matters. He is author of "Conducting scientifically crafted child custody evaluations" (Sage, 1998) and more than 30 articles in the family law area. He is a principal in the national consulting consortium Child Custody Consultants and can be reached through their Website: <Child-Custody-Consultants.com>.

Dana Greenhut, PhD, was a pre-doctoral psychology intern at Hawthorn Center when this study was conducted. She is presently employed as a psychologist at Psychiatric Behavioral Medicine Professionals in Livonia, MI.

Address correspondence to: James N. Bow, PhD, Hawthorn Center, 18471 Haggerty Road, Northville, MI 48167 (E-mail: bow@michigan.gov).

Journal of Child Custody, Vol. 2(4) 2005
Available online at http://www.haworthpress.com/web/JCC

doi:10.1300/J190v02n04_01

ABSTRACT. A survey of experienced psychologists was conducted to analyze the procedures used in administering, scoring, and interpreting the MMPI-2 and MCMI-II/III in child custody evaluations. Findings revealed some major concerns, such as over-reliance on computer-generated interpretive reports, lack of knowledge about base rate cut-offs for interpreting the MCMI-II/III, and a failure to consider available MMPI-2 context specific normative data for child custody litigants. Also, a smaller group of psychologists hand scored the protocols or failed to verify computer keypad data entry, and allowed examinees to take the test in the lobby. The rationales underlying these concerns are outlined, along with the need for psychologists to adhere closely to professional and ethical standards in this regard.

KEYWORDS. MMPI-2, MCMI-II, MCMI-III, Millon, psychological testing, personality assessment, custody evaluations, and forensic evaluations

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and Millon Clinical Multiaxial Inventory-II/III (MCMI-II/III) are the most widely used tests in child custody evaluations (Quinnell & Bow, 2001). The MMPI-2, used by over 90% of child custody evaluators (Ackerman & Ackerman, 1997; Bow & Boxer, 2003; Hagen & Castagna, 2001; Quinnell & Bow, 2001), has a solid empirical research base and includes context specific normative data for child custody litigants (Bagby, Nicholson, Buis, Radnovik, & Fidler, 1999; Bathurst, Gottfried, & Gottfried, 1997; Flens, 2004; Strong, Greene, Hoppe, Johnston, & Olesen, 1999). The MCMI-III's popularity has increased over the past ten years. Ackerman and Ackerman (1997) found that 34% of psychologists used the MCMI-II/III in child custody evaluations; however, this number exceeded the 50% range in more recent studies done by Quinnell and Bow (2001) and Bow, Quinnell, Zaroff, and Assemany (2002). Context specific normative data for child custody litigants have been established for the MCMI-III as well (Flens, 2004; McCann, Flens, Campagna, Collman, Lazzaro, & Connor, 2001). Given their high usage rates in child custody cases these tests will be subject to increased legal scrutiny, particularly with the adoption of Daubert (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993) by most

states (Hamilton, 1998), and other similar admissibility standards (e.g., McDaniel v. CSX Transportation, Inc., 955 S.W.2d 257, Tenn. 1997; Ramirez v. State, 651 So.2d 1164, Fla. 1995).

This article will focus on the use of the MMPI-2 and MCMI-II/III in child custody cases. A detailed explanation of the development, administration, scoring, and interpretation of these tests will be provided for the multidisciplinary readership. A survey of psychologists' practices in this area will be discussed as well.

The original Minnesota Multiphasic Inventory (MMPI), published in 1943, is based on an empirical keying approach. In this approach, responses to individual test items are treated as unknowns, and empirical item analysis is used to identify test items that differentiate between criterion groups. The MMPI includes validity scales to detect deviant test-taking attitudes, and a correction (K) scale that was later added to identify clinical defensiveness. The latter scale was also used as a correction factor for clinical scales; however, for over fifty years researchers have questioned the validity of using the K-correction on the MMPI and MMPI-2 (Bartlow, Graham, Ben-Porath, Tellegen, & McNulty, 2002; Comrey, 1958; Graham, 2000; Greene, 2000; Hunt, 1948; Schmidt, 1948).

Over the years, more than 10,000 studies have been published on the MMPI (Graham, 2000). Research found the MMPI to be an effective tool for generating descriptions and inferences about individuals based on their profiles. As a result, it became the most widely used personality test in the United States (Graham, 2000).

A few decades after its publication, concerns were raised about the MMPI's original standardization sample (i.e., convenience sample of exclusively Caucasian individuals from the Minneapolis, Minnesota, area, with about eight years of formal education), problems with item content (i.e., outdated terms, poor grammar, and sexist language), and limited range of questions in the item pool. In response to these concerns, a restandardization project was initiated to revise item content/pool and to collect contemporary norms reflective of the general population. In 1989, the MMPI was replaced with a newer version (e.g., MMPI-2).

The final version of the MMPI-2 consisted of 567 items, one more than the original version. Sixteen repeat items were deleted and 14% of the items were reworded. The four main validity scales (Lie, Infrequency, K-correction, and Cannot Say) were retained and three additional validity scales (Back Page F, Variable Response Inconsistency, and True Response Inconsistency) were added. The ten standard scales

remained intact, along with the Harris-Lingoes subscales. Fifteen new content scales were included, along with content component scales. The subscales and content scales provide further refinement of major themes and dimensions on the MMPI-2.

Instead of using linear T-scores that maintain the same distribution as the raw scores, but do not have the same meaning for every scale (i.e., different percentile ranks), the MMPI-2 uses uniform T-scores that allow a given T-score to be equivalent with regard to percentile rank across clinical scales. The critical cut-off was designated as a T-score of 65 rather than the previous T-score of 70. This creates a uniform standard for comparison.

The normative sample was updated to include 2,600 adults (1,138 men and 1,462 females). Demographically the group mirrored the 1980 census but was more highly educated.

An abbreviated version of the MMPI-2 is available. This version consists of the first 370 items of the original test and provides only the main validity and standard scales. The content and supplementary scales are not included. Consequently, the abbreviated form of the MMPI-2 is not recommended (Graham, 2000). Furthermore, the full version has a voluminous amount of research to support its use, which is critical in a forensic setting.

In a forensic setting, Otto (2002) stresses the importance of using the MMPI-2 to generate hypotheses that will lead to further inquiry. He also notes the value of the MMPI-2 in assessing mental status and functioning, identifying behavioral patterns and personality styles that may adversely impact parenting and co-parenting, and analyzing response style (e.g., defensiveness, symptom exaggeration, etc.).

The MMPI-2 manual (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) describes administration, scoring, and interpretation of the test. According to the manual, the typical testing situation requires a table, chair, and an environment free from distraction. Administration may take place in a waiting room, but only if the examiner can ensure that others will not bother the examinee. However, Pope, Butcher, and Seelen (2000) discouraged administration of the MMPI-2 in a waiting room area for forensic purposes, noting an unmonitored setting can cause invalid and misleading findings. There is no way to ensure that the examinee was not assisted or distracted by others.

The test booklet and answer sheet format is the most common method for administering the MMPI-2. This method requires hand scoring or computing scoring. The test may also be administered by

computer. The computer program has the capability of scoring the examinee's responses and providing a report.

The manual states that an eighth grade reading level is necessary for comprehension of the content of all MMPI-2 items (Butcher et al., 1989, 2001). If the adequacy of an examinee's reading level is questionable, the manual states that a reading test should be given. When an examinee lacks adequate reading skills, an audiocassette version of the test is available through Pearson Assessment, which insures standard administration.

Two approved computer-scoring services are available: Pearson Assessment (formerly NCS Assessment) and Caldwell Report. Pearson Assessment offers the Microtest Q Assessment System Software, which allows scanner or keypad data entry. The keypad data entry method provides a way to verify that all responses were accurately entered. This is called the "verification" process and requires the examiner to enter the data twice to ensure accuracy. When the examinee takes the paper and pencil version of the MMPI-2, a mail-in service is also available. A variety of report options are offered, including Extended Score Report, Adult Clinical Interpretive Report-Revised, and Forensic Interpretive Report, which includes child custody as one of the six forensic interpretive reports.

The Caldwell Report offers a mail-in or modem service that provides two report options: interpretive report and child custody report. Greene and colleagues also provide an interpretive report for scored MMPI-2 protocols through Psychological Assessment Resources. Currently, the use of interpretive reports has been called into question, with particular concerns focused on the derivation of interpretive statements found in the computer generated interpretive reports (Behnke, 2004; Flens, 2004; Otto, 2002).

The MCMI-II/III is an objective, multi-scale personality inventory that assesses personality syndromes and disorders. It is based on a comprehensive theory of personality and psychopathology developed by Millon, and corresponds closely with the diagnostic system in the Diagnostic and Statistical Manual published by the American Psychiatric Association. It is important to note that the MCMI is a criterion-referenced test rather than a norm-referenced test. In other words, it examines the probability that a person has the presence or prominence of a particular trait, syndrome, or disorder (Millon et al., 1997, pp. 113-114) rather than comparing a person's deviance with that of a standardized sample. As a result, the Millon scales use base rates (BR) rather than T-scores. Base rates range from zero (0) to 115, with a median of 60.

Scores from 75 to 84 indicate the presence of a symptom/trait, while scores at or above 85 indicate a prominence of a symptom/trait/disorder.

The Millon scales have undergone several revisions since the publication of the original version in the early 1980s. A revised version (MCMI-II) was developed in the late 1980s to improve the test's consistency with the Diagnostic and Statistical Manual of Mental Disorders-Third Edition Revised (DSM-III-R; American Psychiatric Association, 1987). The current version of the Millon, the MCMI-III, was published in 1994. It corresponds more closely than earlier editions to the most recent DSM version (DSM-IV; American Psychiatric Association, 1994, 2000). The MCMI-III has 175 items, of which 95 items were not included in the previous version (MCMI-II). As a result, research involving the MCMI-II cannot be generalized to the MCMI-III.

The MCMI-III consists of four validity scales, 14 personality pattern scales, and 10 clinical syndromes. The normative sample involves 998 males and females with a variety of clinical diagnoses. This was a clinical sample of convenience or sample of opportunity, rather than a census matched sample.

Traditional administration of the MCMI-II/III involves a pencil and paper format. Minimum reading level for the MCMI-III is eighth grade. Audiotape and computer administration is also available for those to whom the test items need to be read. According to the manual, the optimal testing setting should be reasonably comfortable and free of distractions. The manual states the test should never be taken home to complete.

The MCMI-II/III can be scored by hand, computer, or mail-in service. Pearson Assessment offers mail-in service and provides a computer assessment software program, called Microtest-Q, for keypad entry or scanning. As with the MMPI-2, the keypad entry method allows for verification of the data. Profile and interpretive reports are available. Also, Psychological Assessment Resources provides an interpretive software report program.

With the wide use of the MMPI-2 and MCMI-II/III in child custody work, it is important to understand psychologists' procedures in administering, scoring, and interpreting these tests. It has been the experience of the first three authors that many problems are found in these areas when they are requested to perform work product reviews. However, there is no formal research in this area. Consequently, the present study was developed to formally address this issue by exploring the contrast between what is done in daily practice and what is recommended in the

manual/literature by surveying a national sample of psychologists actively involved in child custody practice. It is hoped this information will help clarify the forensic application of these tests in child custody work.

METHOD

Names of potential survey participants were obtained from four sources: (1) an Internet search of psychologists involved in child custody work; (2) child custody evaluators belonging to a child custody listserv; (3) a public domain referral list of members of the American Board of Forensic Psychology; and (4) individuals known to the authors through conferences, publications, and professional activities.

A cover letter, informed consent sheet, blank survey, and stamped self-return envelope were sent to all potential participants. These psychologists were requested to anonymously complete and return the eight-page survey, which included a portion focusing on administering, scoring, and interpreting the MMPI-2 and MCMI-II/III in child custody evaluations. Participants were informed that all data would be coded and analyzed on a group basis as a means of protecting confidentiality. Those interested in the group findings were asked to contact the first author by e-mail. Approximately one month after the initial mailing, a reminder letter was sent.

Three hundred surveys were mailed to potential participants, 135 of which were returned. Of those, 89 respondents met the selection criteria, i.e., completed surveys from psychologists who use testing during CCEs. Of the other returned surveys, three refused to complete the survey, nine reported that they no longer conduct CCEs, five did not use testing in CCEs, two retired from the field, three were non-psychologists, and 24 envelopes were undeliverable. Overall, the adjusted return rate was 35%, which compares very well with other child custody survey research. For example, the Keilin and Bloom (1986) and Ackerman and Ackerman (1997) studies had return rates of 27% and 25%, respectively. Nevertheless, the relative small number of respondents in this study limits the ability to generalize the findings.

Seventy percent of the study respondents were male, and almost all were Caucasian (95.4%). The average age was 54.13 (S.D. 6.53). Ninety-eight percent held doctoral degrees and 21% were diplomates from the American Board of Professional Psychology. The most common diplomate was in forensic psychology (71%). Most respondents

were originally trained as clinical psychologists (67%), followed by counseling psychology (18%), school psychology (5%), forensic psychology (2%), and other (8%).

The respondents averaged 21.88 (S.D. 7.54) years of experience, with 17.81 (S.D. 7.56) and 16.50 (S.D. 7.00) years in the forensic and child custody fields, respectively. Thirty-three percent of their practice was devoted to child custody work. The estimated mean number of evaluations over the course of their career was 312, with an estimated mean of 14 evaluations per year. Ninety-three percent were in private practice, with almost all working in an urban/suburban area (91%). Respondents were from 33 states, with the following geographic distribution: 29.5% from the Midwest, 29.5% from the West, 24% from the South, and 17% from the East.

RESULTS

MMPI-2

Of the 89 respondents, 87 used the MMPI-2 (97.8%) in child custody evaluations. These respondents reported a mean of 20.27 years (S.D. 7.75) of experience with the MMPI/MMPI-2 and had administered an estimated average of 377 MMPI/MMPI-2s in child custody cases over the course of their careers; the estimated range was from 18 to 2,500. The vast majority of respondents administered the test in an office setting, but approximately 25% of respondents had it completed in the lobby. None reported giving the test to examinees to complete at home.

Regarding the required reading level for the MMPI-2, respondents most commonly indicated an 8th grade level (32.65%), with a mean reading level of 7.36 (S.D. 1.52). However, 48% of respondents thought the reading level was below the 8th grade. Respondents indicated that they most frequently evaluate reading level by educational background and informal assessment, i.e., examinee reading a few items aloud (29%). Otherwise, the most common methods were educational background alone (22%); followed by educational history in combination with formal assessment (e.g., administering a reading test) and informal assessment (15%); and finally, educational background and formal assessment (14%).

Respondents indicated that they almost always (98%) administer the standard test comprised of 567 items. Seventy-nine percent of respondents reported that they administered the MMPI-2 using standard in-

structions, while the remaining respondents modified the instructions, i.e., information added to standard instructions. Administration occurred by paper and pencil in almost all cases, with only 10% using a computer.

In terms of scoring the paper and pencil administration, 49% of the respondents entered the data via computer keypad, 32% used a mail-in service, 10% hand scored the protocol, and 9% used a scanner. Of those respondents entering the data via computer keypad, only 65% verified their data entry.

Of those respondents (N = 78) that reported using a scoring service, the overwhelming majority (83.1%) used Pearson Assessment (formerly NCS Assessments), 13% used Alex Caldwell's service, and the remaining respondents indicating "other." The following types of scoring reports from NCS-Pearson were used by respondents: 46.2% Extended Scoring Report, 30.8% Adult Interpretive Report, 16.7% Forensic or CCE Interpretive Report, and 6.4% Basic Scoring Report. For interpretive reports, 61.4% used the Pearson Interpretive Report by James Butcher, 25% used the Caldwell Report, and 13.6% used Roger Greene's Adult Interpretive System by Psychological Assessment Resources.

In terms of using the information from the interpretive report in a child custody report, respondents (N = 46) reported the following: 58.7% indicated that they used the interpretive report to write their own description (i.e., used information from the report to write their own summary of the findings), 21.7% indicated they inserted statements or paragraphs with slight modifications from the report, 10.9% reported they only used scores and wrote their own description based on information from other sources, and 8.7% inserted sentences or paragraphs verbatim from the report. None reported inserting the complete interpretive report verbatim.

Those respondents not using interpretive reports were queried about their reason. Thirty-nine percent of respondents thought they were knowledgeable enough to interpret the scores without an interpretive report and 29% of respondents expressed concern about how interpretive statements were generated and the validity of the program. Other reasons given by respondents were attorneys might gain access to the report (15.8%), the high cost of interpretive reports (13.2%), and the need to use context specific normative data based on child custody litigants in the interpretive process (5.3%).

Twenty percent of respondents reported that they also hand scored some supplementary scales. Half of this group indicated they hand

scored the following validity scales: Wiggins Social Desirability Scale, Positive Malingering, and Edwards Social Desirability Scale.

Respondents were asked to rate on a Likert scale (1-Not Useful to 7-Extremely Useful) the usefulness of the MMPI-2 validity, clinical, and supplementary scales in assessing parental test behavior in child custody evaluations (see Table 1). The Lie, K-correction, and Infrequency (F) scales all received mean ratings above 5.0. In addition, when respondents were asked separately to identify the most useful validity scales, the highest percentage (22.5%) identified this triad of scales. When respondents were asked if they would re-administer the MMPI-2 when a "fake good" profile was obtained, with instruction to "be more honest," only 14% indicated they would do so.

As shown in Table 1, six of the ten clinical (standard) scales had mean ratings of 5.0 or higher, with the top three being Psychopathic Deviant, Depression, and Schizophrenia. The scale receiving the lowest mean rating was Masculinity-Femininity.

Forty-four percent of respondents indicated they typically interpret T-scores below the recommended cut-off of 65. The most commonly mentioned adjusted cut-off used by respondents was 60, reported by 63% of respondents. Also, respondents indicated that 51% used non-K-corrected scores in their analysis of MMPI-2 data, 37% utilized the new MMPI-2 Restructured Clinical scales, 22% consider the non-gendered T-scores, and only 9% plotted the MMPI-2 data on the original MMPI profile.

Content and supplementary scales were interpreted by 79% and 74% of respondents, respectively. The majority of respondents (62%) reported they only interpret such scales when the particular basic scale is elevated. Table 1 shows the mean ratings respondents gave the supplementary scales in regards to their usefulness in understanding parental behavior. Almost all mean ratings were 4s, which indicates only moderate usefulness. The top three were behavioral dyscontrol scales (MacAndrew, Addiction Admission, & Hostility). The two lowest mean ratings involved gender roles (Gender Role-Masculine & Gender Role-Feminine), which is congruent with the lowest mean rating (Masculinity-Femininity) on the Standard (clinical) scales.

In terms of subscales, 59% of respondents reported that they typically analyze and interpret the Harris-Lingoes Subscales, but fewer respondents (42%) reported doing so with the Content Component scales. Almost 60% indicated they analyze and interpret such scales when the base scale is at or above a T-score of 65. About 20% of respondents indicated they analyzed and interpreted these subscales whenever an ele-

TABLE 1. Usefulness Ratings of MMPI-2 Validity, Clinical, and Supplementary Scales

Type of Scale	Mean	S.D.
Validity		
Lie (L)	5.87	1.18
K-Correction (K)	5.82	1.13
Infrequency (F)	5.34	1.46
Superlative (S)	4.94	1.74
Back Infrequency (F_B)	4.47	1.70
Infrequency-Psychopathology (F_p)	4.21	1.62
Variable Response Inconsistency (VRIN)	4.21	1.82
True Response Inconsistency (TRIN)	4.15	1.79
Unanswered Items (?)	4.13	1.98
Clinical		
Psychopathic Deviant	5.98	1.20
Depression	5.56	1.24
Schizophrenia	5.55	1.22
Mania	5.49	1.19
Paranoid	5.43	1.31
Psychasthenia	5.07	1.30
Hysteria	4.96	1.34
Hypochondrias	4.51	1.47
Social Introversion	4.33	1.57
Masculinity-Femininity	3.30	1.84
Supplementary		
Mac Andrew-Revised	4.93	1.55
Addiction Admission	4.75	1.65
Hostility	4.63	1.58
Anxiety	4.61	1.42
Ego Strength	4.60	1.51
Addiction Potential	4.51	1.72
Dominance	4.45	1.60
Overcontrolled Hostility	4.37	1.77
Repression	4.35	1.44
Social Responsibility	4.27	1.60
Post Traumatic Stress Disorder	4.12	1.61
Gender Role-Masculine	2.85	1.53
Gender Role-Feminine	2.84	1.53

Note. Each scale was rated on a Likert scale (1-Useless to 7-Highly Useful).

vation occurs on the subscale, regardless of the score on the main scale, which is contrary to interpretative guidelines.

Only 55% of respondents indicated that they used context specific normative data in interpreting MMPI-2 child custody scores. The most commonly cited normative data were Bathurst, Gottfried, and Gottfried (1997), reported by 42% of respondents, and followed by respondents using a combination of normative data from different researchers. The overwhelming majority of these respondents (85%) indicated that they compared the context specific normative data to the regular MMPI-2 norms. Nine percent reported that they used the context specific normative data to create alternative explanations.

When respondents were asked how long they typically wait to re-administer the MMPI-2 in a follow-up evaluation, the majority (51%) indicated they wait 6-12 months, while 31% reported waiting 12-18 months. Seven percent of respondents indicated they waited less than six months, with a corresponding percentage waiting 18-24 months. Two percent of respondents reported waiting over 24 months and one respondent reported never retesting.

MCMI-II/III

Sixty-three percent of respondents (N = 56) reported using the MCMI-II/III in child custody evaluations. Respondents reported a mean of 12.42 years of experience with the MCMI scales. The estimated mean number of MCMIs administered by respondents in their child custody career was 211.

In terms of the required educational level for taking the MCMI-II/III, the most common response was 6th grade (38%), with 54.5% of respondents indicating a reading level somewhere below the recommended reading level (8th grade); only 31% identified the recommended reading level. Respondents most frequently evaluated reading level based on educational background and informal assessment, i.e., having examinees read a few items aloud (39%). Otherwise, the most common methods were educational background alone (26%); followed by educational background in combination with formal assessment (i.e., administering a reading test) and informal assessment (16%); and finally, educational background and formal assessment (14%).

The administration site for the MCMI-II/III was usually an office setting (72%), with the remainder of respondents using the lobby of their office. None of the respondents reported allowing the examinee to take

the test home. The overwhelming majority of respondents used the standard instructions (93%), with only a small percentage modifying the instructions (i.e., adding information to the instructions). Eight-five percent of respondents administered the test by paper and pencil, with the remaining respondents using computer administration.

For the paper and pencil format, the most common scoring system involved the examiners entering the data via computer keypad (56%), followed by mail-in scoring service (32%), hand scoring (8%), and scanner (4%). Of those respondents entering the data via computer keypad, the overwhelming majority (85%) verified the data entry (i.e., reentering all the data a second time). Of those using a scoring service, almost all used Pearson Assessment (94%), and 62% of these respondents obtained the interpretive report, with the remainder using the profile report.

Regarding respondents' use of the interpretive report in their child custody reports, 64% indicated they wrote their own description based on the findings from the MCMI-II/III report, 14% reported inserting sentences or paragraphs from the MCMI-II/III report with slight modifications, 14% indicated they only used the scores and wrote their own description based on information from other sources, and 8% reported inserting sentences or paragraphs in verbatim form from the MCMI-II/III report.

Respondents who did not use interpretive reports were queried as to the reason. The most common response was that they did not know how the interpretive statements were generated (33%), the perception that they were sufficiently knowledgeable to interpret the scores without an interpretive report (33%), the high cost of the interpretive report (17%), and concern that the interpretive report might over-pathologize examinees (17%).

Respondents were asked what cut-off score they use for interpreting base rate scores on the MCMI-II/III. Sixty percent of respondents correctly identified a base rate of 75, with 17% and 15% of respondents indicating base rates of 70 and 65, respectively. Four percent of respondents reported a base rate of 60, and the same percentage indicated a base rate of 80. It is also important to note that only a slight majority of respondents were able to accurately identify the cut-offs for the presence of a trait/syndrome (BR = 75; 57.1%) and prominence of a trait/syndrome (BR = 85; 52.9%).

Forty-six percent of respondents indicated they modified the cut-off score for child custody evaluations. Of those respondents, 76% reported they did so because of child custody research involving the MCMI-II/

III. Another 9.5% of respondents reported concerns about the MCMI-II/III over-pathologizing examinees, while the same percentage of respondents indicated concerns about gender bias. One respondent also reported using non-corrected scores to address the latter issue. When respondents were directly asked if they had concerns about gender bias in the MCMI-II/III, 42% indicated yes; the vast majority of this entire group (94%) indicated their awareness of current research on this issue.

Respondents were asked to rate the usefulness of MCMI-II/III scales (Clinical Personality Patterns, Severe Personality Pathology, Modifying Indices, Severe Clinical Syndromes, and Clinical Syndromes) in child custody evaluations using a Likert scale (1-Not Useful to 9-Extremely Useful). All scales received mean ratings of 6, indicating a moderate level of usefulness.

Ninety-one percent of respondents indicated they would not re-administer the MCMI-II/III if they obtained a "fake good" validity profile with instructions to "be more honest."

DISCUSSION

The usage rate of the MMPI-2 and MCMI-II/III in this study was consistent with findings reported in past child custody research (Ackerman & Ackerman, 1997; Bow et al., 2002; Hagen & Castagna, 2001; Quinnell & Bow, 2001). These tests are both widely used in the child custody field. Therefore, it is imperative to understand the procedures psychologists use in administering, scoring, and interpreting these tests, especially in light of the increased legal scrutiny provided by the Daubert standard. This study was developed to address these issues, and hopefully help psychologists and the court better understand this area.

The findings from this study revealed some grave concerns about the administration, scoring, and interpretation of the MMPI-2, which have serious implications for professional practice, as well as child custody practice. Almost half of the respondents underestimated the required reading level (8th grade) for the MMPI-2 and MCMI-II/III, which might invalidate the results for those with inadequate reading skills. Also, approximately 25% of respondents allowed examinees to take the tests in the lobby, which is ill-advised. As noted by Pope et al. (2000), testing cannot be adequately monitored in such a setting. There is no way of ensuring that the examinee was not assisted by others in completing the test or that the setting was free from distraction. Concerning

scoring, approximately 10% of respondents hand scored the MMPI-2 and MCMI-II/III. Hand scoring is simpler for the MMPI-2; for the MCMI-II/III, it is extremely difficult (Millon, 1997). The probability of error increases significantly with hand scoring (Allard & Faust, 2000; Simons, Goddard, & Patton, 2002); consequently, it is an unwise practice in forensic work. Scoring problems may also occur while examiners are entering the data via computer keypad. Almost half of the respondents used this method, but only 65% of them verified their MMPI-2 data entry. MCMI-II/III data was verified at a higher rate (85%). This discrepancy in verification rates between the two tests is probably a function of the increased time involved for re-entering MMPI-2 data. Nevertheless, it is critical that all keypad data entry be verified; errors can create major changes in the profile and interpretation, which might negatively impact the general findings and conclusions.

Another huge concern was the wide use of computer generated interpretative reports. For the MMPI-2, 47.5% of respondents used such reports; whereas 62% of respondents used them with the MCMI-II/III. This finding raises six major issues. First, psychologists lack knowledge of the underlying algorithms or decision rules used for the computer-generated interpretive statements (Otto, 2002). Evidentiary rules require that experts identify the basis of their opinion. Further, ethical guidelines require that psychologists retain responsibility for the appropriate application of automated reports (Behnke, 2004). This concern is highlighted by the fact that 22-30% of respondents acknowledged inserting the exact sentences or slightly modified sentences from the computer generated interpretive report into their reports.

Second, Flens (2004) contends that the following questions should be answered prior to using such programs: (a) Is the program an actuarial interpretation program or an automated interpretation? (b) What is the level of significance regarding test scores? (c) Are there different levels of significance for different scales? (d) At what point does the program actually generate a statement for a particular scale? (e) Are there different statements depending on the level of elevation for any given scale? (f) Does the program consider profile configuration or combination of scales, or are the statements based on a single-scale elevation? (g) Does the program consider the response style when offering the statements? (h) Does the program use context-specific normative data to facilitate interpretive statements? Flens points out that most of these questions are unanswerable because they are proprietary secrets. Psychologists are placed in an indefensible position because they are

unable to answer questions due to the lack of peer reviewed information about the development and validation of the interpretative statements.

Third, psychologists need a solid knowledge of test theory and psychological assessment; computer-generated interpreted reports cannot act as a substitute (Otto, 2002). Further, psychologists must ensure that they are competent to use the tests and render interpretations that the service has provided; psychologists must stay within their boundaries of competency (Behnke, 2004). The MCMI-II/III is a complex and difficult test to understand and interpret, but its main attraction is that it is relatively quick to administer. This might result in psychologists' using it in child custody evaluations without the prerequisite knowledge/skills.

Fourth, findings from the MMPI-2 and MCMI-II/III are affected by the setting (e.g., clinical vs. forensic) and context (e.g., child custody vs. personal injury) of the administration. These factors need to be considered in the interpretation, which most computer-generated interpretive reports fail to do. Within the last few years, Pearson Assessment and Caldwell have developed MMPI-2 computer-generated forensic interpretive reports, with a particular emphasis on child custody. However, they suffer from many of the same above-mentioned concerns.

Fifth, the MCMI-II/III computer-generated interpretive reports do not consider response style or gender bias issues in child custody cases (Flens, 2004). Recent research has indicated serious problems in these areas (Hynan, 2004; McCann et al., 2001). Lastly, MCMI-II/III computer-generated interpretive reports overstate psychopathology. McCann (2002) recommends that professionals avoid relying on MCMI-III computer-generated interpretive reports.

It is noteworthy that respondents who did not use computer-generated interpretive reports were able to identify many of the above-mentioned concerns as the reason(s) for not using them. Therefore, some psychologists are aware of these critical issues and avoid such interpretive reports. It is hoped that all psychologists become increasingly aware of these issues and should not rely on interpretative reports without knowing where each interpretative statement was derived.

It is also worth mentioning that evaluators that self-interpret the MMPI-2 or MCMI-II/III must have their own interpretive decision rules that are based on empirical research, which can be logically applied and clearly articulated during testimony.

Of extreme concern was the finding that almost half of the respondents were unaware of the MCMI-II/III cut-offs for the presence (BR 75-84) and prominence (BR $\geq$ 85) of a trait/syndrome. This lack of ba-

sic knowledge raises serious practice and ethical issues, particularly regarding the competency of these examiners to use the test. This might be due to their over-reliance on interpretative reports or an inability to understand base rates. Nevertheless, we find this lack of knowledge of great concern because it could lead to erroneous findings and conclusions. In addition, the use of a test that a psychologist does not know how to use and interpret may suggest that one is practicing outside his/her competency, a violation of ethical standards.

Only 55% of respondents indicated they used MMPI-2 context specific normative data with child custody litigants. This is surprising considering the availability of such norms and the numerous publications over the past ten years regarding this issue (Bagby et al., 1999; Bathurst et al., 1997; Butcher, 1997; Posthuma & Harper, 1998; Schneck, 1996; Siegel, 1996; Strong et al., 1999).

In regards to interpreting MMPI-2 scales, the typical cut-off is 65 or higher (Butcher & Williams, 2000; Graham, 2000; Greene, 2000). Some debate has focused on interpreting scores below 65 for defensive profiles. Graham (2000) and Butcher (1992) recommended a T-score cut-off of 60 or higher for such profiles. About half of the respondents in this study concurred with this recommendation.

The most useful MMPI-2 validity scales were seen by respondents as Lie, K-corrected, and Infrequency (F). The Lie and K-scales are often cited in the child custody literature as being the most elevated validity scales (Bagby et al., 1999; Siegel, 1996; Strong et al., 1999). Among clinical scales, Scales 4, 2 and 7 (the Psychopathic Deviant, Depression and Schizophrenia scales, respectively) were identified by respondents as being most useful in identifying concerns about parenting behavior. This is expected considering the impact these traits and symptoms might have on parenting. Scale 4 (i.e., Psychopathic Deviant scale) is one of the most commonly elevated scales in child custody cases (Caldwell, 1997) and among individuals receiving marital counseling or reporting marital problems (Otto & Collins, 1995). It is often associated with family conflict and anger. The high frequency of this elevation may also be due to issues associated with the K-correction, which is used to correct for defensiveness. Scale 4 receives a correction of forty-percent from the K Scale. Consequently, when the K Scale is elevated due to a favorable response style, which commonly occurs in child custody cases, it elevates the score on Scale 4. Therefore, this issue must be considered in the interpretation of the scale.

The overwhelming majority of respondents indicated that they interpret MMPI-2 content and supplementary scales, although they rated

them as only moderately useful. The latter is probably due to the common "fake good" response style seen in child custody cases, which results in lower profiles, with few, if any, significant elevations on these scales. Common elevations on the Supplementary Scales include Repression (R), Dominance (Do) and Social Responsibility (RE) (Flens, 2004). Further, some of the supplementary scales lack solid empirical support and should be avoided altogether in child custody cases, particularly the Overcontrolled Hostility scale (hereinafter O-H). This scale was originally constructed to differentiate assaultive prisoners from those without such histories. This scale lost three items in the re-standardization of the MMPI-2, and little research currently exists for using this scale outside prison settings or as a predictive measure of assaultive behavior. Bathurst et al. (1997), for example, found that both male and female custody examinees had elevated O-H scores of approximately one standard deviation above the mean (i.e., T = 60). An examination of O-H items suggests they are worded in such a way that examinees hoping to portray a favorable impression will receive an elevated score (Flens, 2004), a common finding on the MMPI-2 in this context. Consequently, its use with a child custody population is inappropriate. It is disconcerting that respondents in this study rated it as moderately useful.

Only 42% of respondents reported using the MMPI-2 content component scales. However, 20% of that group indicated that they analyzed and interpreted the corresponding subscale whenever it was elevated, regardless of the score on the main scale. This is contrary to recommended practice; several authors report that both content component scales and Harris-Lingoes subscales should be used to refine the interpretation of elevated content scales (Butcher et al., 2001; Butcher & Williams, 2000; Graham, 2000; Greene, 2000).

CONCLUSION

This study was developed to explore the administration, scoring, and interpretation of the MMPI-2 and MCMI-II/III in child custody evaluations. Findings from the survey revealed the following major concerns. First, we found that respondents tend to over-rely on the use of computer-generated interpretive reports in reaching conclusions about an individual's functioning rather than using the interpretive reports as hypotheses generated from an alternative third party. This was particularly problematic with respondents' reliance on the MCMI II/III interpretive report, which tends to overstate psychopathology and fails to consider

gender bias and custody response-style issues. Second, almost half of the respondents were not aware of the base rate cut-offs for interpreting the MCMI-II/III, raising important questions about evaluators' understanding of how to interpret accurately MCMI-II/III test results. This lack of knowledge might be due to respondents' over-reliance on interpretations offered in computer-generated interpretive reports and/or inadequate training with the test. Third, only half of the respondents used context specific normative data when interpreting the meaning of MMPI-2 scores. This finding was surprising in light of the almost decade old data addressing MMPI-2 context specific norms for male and female custody litigants (Bathurst et al., 1997). Fourth, we found troubling the continued use by a small group of respondents of hand scoring the tests. More troubling was the finding that a larger proportion of respondents did not verify their computer keypad data entry to ensure the accuracy of their original data entry. Fifth, about one-quarter of respondents allowed the tests to be completed in the lobby of their office. This practice is ill-advised, particularly considering that forensic evaluators must ensure the integrity of the test administration. Lastly, almost half of the respondents underestimated the reading level for the MMPI-2 and MCMI-II/III. It is important for psychologists to understand the reading skills necessary to administer these tests; otherwise they will jeopardize the validity and reliability of the instruments by administering the tests to people educationally unprepared to take the tests.

Overall, this study raises several important concerns about psychologists' understanding of how to properly use and interpret test data from the MMPI-2 and MCMI-II/III. We are concerned that in many cases, psychologists' lack of knowledge about proper interpretation may reveal a lack of competence in the use of such tests. This is particularly surprising considering that forensic work is often closely scrutinized by others, and it would be expected that forensic evaluators would be more knowledgeable about the administration, scoring, and interpretation of tests used in the evaluation process than evaluators practicing in non-legal settings. The lack of compliance with testing standards and psychological ethics is quite troubling. Furthermore, the negative impact of inappropriately administering and scoring tests and the misinterpretation of test results upon the evaluation and legal process are problematic. It is hoped these findings will create awareness of these testing difficulties and promote further education of psychologists.

REFERENCES

Ackerman, M. J., & Ackerman, M. (1997). Custody evaluation practice: A survey of experienced professionals (revisited). *Professional Psychology: Research and Practice*, *28*, 137-145.

Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment*, *7*, 119-129.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed rev.). Washington, DC: Author.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th edition-text revision). Washington, DC: Author.

Bagby, M., Nicholson, R., Buis, R., Radnovik, H., & Fidler, B. (1999). Defensive responding on the MMPI-2 in family custody and access awards. *Psychological Assessment*, *11*, 24-28.

Bartlow, D. L., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & McNulty, J. L. (2002). The appropriateness of the MMPI-2 K correction. *Assessment*, *9*(3), 219-229.

Bathurst, K., Gottfried, A., & Gottfried, A. (1997). Normative data for the MMPI-2 in child custody litigation. *Psychological Assessment*, *9*, 205-211.

Behnke, S. (2004). Test-scoring and interpretation services. *Monitor on Psychology*, *35*, 58-59.

Bow, J. N., & Boxer, P. (2003). Assessing allegations of domestic violence in child custody evaluations. *Journal of Interpersonal Violence*, *18*, 1394-1410.

Bow, J. N., & Quinnell, F. A. (2001). Psychologists' current practices and procedures in child custody evaluations: Five years after the American Psychological Association guidelines. *Professional Psychology: Research and Practice*, *32*, 261-268.

Bow, J. N., Quinnell, F. A., Zaroff, M., & Assemany, A. (2002). Assessment of sexual abuse allegations in child custody cases. *Professional Psychology: Research and Practice*, *33*, 566-575.

Butcher, J. N. (1992). Interpreting defensive profiles. In J. N. Butcher & J. R. Graham (Eds.), *Topics in MMPI-2 & MMPI-A interpretation (No. 8)*. Minneapolis, MN: MMPI-2 Workshops and Symposia.

Butcher, J. N. (1997). Frequency of MMPI-2 scores in forensic evaluations. Paper presented at the 32nd Annual Symposium on Recent Developments in the Use of the MMPI-2, Minneapolis, MN.

Butcher, J. N., & Williams, C. L. (2000). *Essentials of MMPI-2 and MMPI-A Interpretation* (2nd Ed). Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., Kaemmer, B. (1989). *MMPI-2-Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation* (rev. ed.). University of Minnesota Press: Minnesota.

Caldwell, A. (1997). *Forensic questions and answers on the MMPI/MMPI-2*. Los Angeles, CA: Caldwell Report.

Comrey, A. L. (1958). A factor analysis of items on the K scale of the MMPI. *Educational and Psychological Measurement*, *18*, 633-639.

Daubert v. Merrell Dow Pharmaceuticals, Inc. 509 U.S. 579 (1993).

Dyer, F. J. (1997). Application of the Millon inventories in forensic practice. In T. Millon (Ed.), *The Millon Inventories: Clinical and Personality Assessment* (pp. 124-139). New York: Guilford.

Flens, J. R. (2004, October). *Advanced institute on psychological testing*. Pre-symposium Institute conducted at the fourth annual Association of Family and Conciliation Courts Child Custody Symposium, Nashville, TN.

Gould, J. W. (2005). *Conducting Scientifically Crafted Child Custody Evaluations (2nd Edition)*. Sarasota, FL: Professional Resource Press.

Graham, J. R. (2000). *MMPI-2: Assessing personality and psychopathology*. (3rd ed.). New York: Oxford Press.

Greene, R. L. (2000). *MMPI-2/MMPI: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

Hagen, M. A., & Castagna, N. (2001). The real numbers: Psychological testing in child custody evaluations. *Professional Psychology: Research and Practice*, *32*, 269-271.

Hamilton, H. G. (1998). The movement from Frye to Daubert: Where do the states stand? *38 Jurimetrics Journal*, 201-213.

Hunt, H. F. (1948). The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *Journal of Consulting Psychology*, *12*, 396-402.

Hynan, D. (2004). Unsupported gender differences on some personality disorder scales of the Millon Clinical Multiaxial Inventory-III. *Professional Psychology: Research and Practice*, *35*, 105-110.

Keilin, W. G., & Bloom, L. J. (1986). Child custody evaluation practices. A survey of experienced professionals. *Professional Psychology: Research and Practice*, *17*, 338-346.

McCann, J. T. (2002). Guidelines for forensic application of the MCMI-III. *Journal of Forensic Psychology Practice*, *2*, 55-69.

McCann, J. T., Flens, J. R., Campagna, V., Collman, P., Lazzaro, T., and Connor, E. (2001). The MCMI-III in child custody evaluations: A normative study. *Journal of Forensic Psychology Practice*, *1*, 27-44.

McDaniel v. CSX Transportation, Inc., 955 S.W.2d 257 (Tenn. 1997).

Millon, T., Davis, R., & Millon, C. (1997). *Millon Multiphasic Personality Inventory-III: Manual* (2nd ed.). Minneapolis.

Otto, R. K. (2002). Use of the MMPI-2 in forensic settings. *Journal of Forensic Psychology Practice*, *2*, 71-91.

Otto, R. K., & Collins, R. P. (1995). Use of the MMPI-2/MMPI-A in child custody evaluations. In Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI*. (pp. 222-252). Thousand Oaks, CA: Sage.

Pope, K. S., Butcher, J. M., & Seelen, J. (2000). *The MMPI, MMPI-2, & MMPI-A in court: A practical guide for expert witnesses and attorneys*. (2nd Ed.). Washington, DC: American Psychological Association.

Posthuma, A. B., & Harper, J. F. (1998). Comparison of MMPI-2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice*, *29*, 437-443.

Quinnell, F. A., & Bow, J. N. (2001). Psychological tests used in child custody evaluations. *Behavioral Science and the Law*, *19*, 491-501.

Ramirez v State, 651 So.2d 1164 (Fla. 1995).

Schenk, P. W. (1996). MMPI-2 norms for child custody litigants. Georgia Psychological Association.

Schmidt, H. O. (1948). Notes on the Minnesota Multiphasic Personality Inventory: The K-factor. *Journal of Consulting Psychology*, *12*, 337-342.

Siegel, J. C. (1996). Traditional MMPI-2 validity indicators and initial presentation in custody evaluations. *American Journal of Forensic Psychology*, *14*, 55-63.

Simon, R., Goodard, R., & Patton. W. (2002). Hand-scoring error rates in psychological testing. *Assessment*, *9*, 292-300.

Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, N. (1999). Taxometric analysis of impression management and self-deception on the MMPI-2 in child custody litigants. *Journal of Personality Assessment*, *73*, 1-18.

SUBMITTED: April 13, 2005
REVISED: June 7, 2005
ACCEPTED: August 30, 2005