Overview and Current Status of the Millon Clinical Multiaxial Inventory

Robert J. Craig

Illinois School of Professional Psychology

The MCMI (Millon, 1983a, 1987, 1994, 1997a) was developed to operationalize Millon’s (1969/1983b) model of psychopathology and has been revised 2 times over the past 20 years to keep pace with changes in theory as well as the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; American Psychiatric Association, 1994). Since its introduction in 1977, it has become one of the most widely used and researched clinical assessment instruments in history, generating over 500 articles and 6 books. It has been translated into several languages and is being used in cross-cultural research. This article presents a historical overview of the measure and summarizes its current status in the literature. I highlight Millon’s original vision for the test, changes that have been made, and its strengths and limitations. Future refinement of the MCMI is already in the planning stages and includes trait subscales for the personality disorders and linking codetypes to theory-derived interventions.

The Millon Clinical Multiaxial Inventory (MCMI; Millon, 1983a, 1987, 1994, 1997a) has become an important and commonly used assessment instrument by clinical psychologists (Watkins, Campbell, Nieberding, & Hallmark, 1995). Butcher and Rouse (1996) reported that, over the past few years, only the Rorschach (Exner, 1993) and Minnesota Multiphasic Personality Inventory–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) have generated more published research.

There is now a substantial literature base associated with the MCMI, with about 500 published articles appearing since the test’s introduction in 1977 (Craig, 1993a, 1997a). Six books have been published devoted primarily to the MCMI (Choca & Van Denburg, 1997; Craig, 1993a, 1993b; McCann & Dyer, 1996; Millon, 1997b; Retzlaff, 1995) with another soon to be added to the list (Strack, in press). The test has generated many reviews and critiques (Dana & Cantrell, 1988; Greer, 1984; Haladyana, 1992; Hess, 1985; Lanyon, 1984; McCabe, 1984; Reynolds, 1992; Wetzler, 1990; Wetzler & Marlowe, 1992; Widiger, 1985). Chap-
ters on the MCMI are now routinely included in textbooks on psychological assessment (Craig, in press-a, in press-b; Davis & Millon, 1997a, 1997b; Gonclaves, Woodward, & Millon, 1994; Groth-Marnatt, 1997; Lehne, 1994; Millon & Davis, 1995, 1996, 1998). The MCMI has been translated into several languages and is used in cross-cultural research (Jackson, Rudd, Gazis, & Edwards, 1991; Luteijn, 1991; Mortensen & Simonson, 1991; Simonson & Mellegard, 1986; Simonson & Mortenstein, 1991). Clinical information continues to be published in the literature pertaining to such matters as MCMI codetypes by diagnoses (Craig, 1995) and the frequency of high point codes among psychiatric patients (Retzlaff, Ofman, Hyer, & Matheson, 1994). Given this level of interest, it is an appropriate time to provide a brief history and update on the current status of this psychological assessment instrument.

THEORETICAL DEVELOPMENT

The MCMI emanated from Millon’s (1969/1983b, 1981, 1990, 1997a) bioevolutionary theory of personality development. Arguing that the structure of a clinical science of personology consists of four elements (e.g., theory, taxonomy, measurement, and intervention), Millon initially developed and evolved his theory of personality, which generated a classification of personality prototypes and personality disorders. He then developed his instruments to assess personality types and personality disorders that were derived from his model. Other articles in this series describe this model and its evolution to its current status (Choca, this issue; Davis, this issue) so it need not be repeated here. The important point for this article is that the test was derived from theory rather than from “dust-bowl” empiricism.

TEST DEVELOPMENT AND VALIDATION

Millon (1983a, 1987, 1994, 1997a) chose a true–false format for the MCMI, but he is not tied to this methodology. He has experimented with a diagnostic rating approach, where a clinician thoroughly familiar with the patient completes the assessment (Tringone, in press; see also Strack, 1987, 1991), and I am developing an instrument to assess personality disorders using adjective checklist methodology. As with any self-report inventory, the patient is required to know something about his or her personality, behavior, and symptoms and be willing to report it honestly on a test. Those aspects of self that are outside of one’s awareness would need to be tapped by different methodologies (e.g., projective tests and partner ratings).

In developing the MCMI, Millon (1983a, 1987, 1994, 1997a) used Loevinger’s (1957) three-step test construction and validation strategy, whereby validation occurs throughout the test development process. At Stage 1—"theoretical--substanz-
tive—items were constructed based on how well their content corresponded with Millon’s theory. Although many of the items correspond well with criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [DSM–IV]; American Psychiatric Association [APA], 1994), others do not and were guided by theory rather than by official nomenclature. Although Widiger (this issue) laments that MCMI items were not written to correspond to theory in that the items do not only reflect the self–other, pleasure–pain, and active–passive polarities, there are items written to reflect these aspects of theory. For example, on the MCMI–III (Millon, 1994, 1997a) Antisocial scale, Item 17 (“As a teenager, I got into lots of trouble because of bad school behavior”) reflects DSM–IV criteria, whereas Item 7 (“If my family puts pressure on me, I’m likely to feel angry and resist doing what they want”) assesses the active–independent variant of Millon’s theory and has no correlation with DSM–IV criteria. There are other examples of this throughout the test, and this is one reason there are only moderate correlations of MCMI personality disorder scales with similar measures.

For the MCMI–I (Millon, 1983a), an initial 1,100-item pool was constructed and split into two equivalent forms. These forms were then administered to clinical patients and items were retained based on scale intercorrelation and endorsement frequency data. Through this process, now called the stage of *internal–structural* or internal validation, a set of 289 items were ready for *external validation* (the third validation stage). These items were given to 167 clinicians who were asked to give the test and complete a diagnostic form on each patient. This process continued until the final set of 175 items was established and validated with convergent and discriminant validity studies.

For the MCMI–II (Millon, 1987), this three-stage validation process continued with (a) an experimental version of 368 items, which included the addition of two new scales; (b) the introduction of an item-weighting system that gave the “prototype” items more weight in scoring; and (c) validity scales and various adjustments were added and refined, increasing the test to 4 validity scales and 13 personality disorder scales. Keep in mind that the test was continuously refined at each stage of the process.

The MCMI–III (Millon, 1994, 1997a) made further changes. Up to 50% of the items from the MCMI–II were changed and the scales were reduced in length; two new scales were added (PTSD and Depressive), noteworthy response items dealing with eating disorders and child abuse were added but not scored on any scale, and the item-weighting scoring system was changed from a 3-point to a 2-point scale.

Some researchers have lamented about the frequency with which the MCMI has been revised, rendering research from earlier versions of the test somewhat obsolete (Choca et al., 1992). Millon’s (1987, 1994, 1997a) position has always been that his instruments are not cast in stone and he will make revisions when it becomes necessary. Although he has maintained that revisions become necessary
when there are new developments in theory, it is not merely coincidence that his
MCMI revisions have corresponded with revisions in the DSM (APA, 1994).

Millon (1983a, 1987, 1994, 1997a) introduced the base rate (BR) score to the
MCMI, which is a transformed score that selects that point in the distribution of
scores whereby the patient has all of the features of the disorder or syndrome at the
diagnostic level. He used this approach because personality disorders and clinical
syndromes are not normally distributed and, therefore, a $T$ score type of transfor-
mation is not appropriate. BR scores of 85 or greater on the personality scales are
said to be diagnostic. Although this approach does maximize diagnostic effi-
ciency, a problem arises with BR scores between 75 and 84, which are said to re-
fect some of the behaviors and traits that define the disorder but not necessarily at
a diagnostic level. Which traits and which behaviors? There is no way to tell. If one
had content-based subscales, then we could determine which of the many domains
have been endorsed by a particular patient and then be more accurate in describing
this person.

PSYCHOMETRIC CONSIDERATIONS

Factor Analysis

Early versions of the MCMI generated a substantial amount of interest in the under-
lying structure, with more than 20 empirical studies on this subject (Craig, 1997a).
This research has generated some debate. Because Millon’s model and instrument
is essentially nonfactorial, there is considerable item redundancy associated with
the test. Millon believes that such covariation exists in nature, and he is not particu-
larly troubled by this redundancy. However, psychometrists argue that item over-
lap on a test produces factorial invariance that is the result of a forced similarity
between scores that artificially elevates the correlations between the scales
(Gibertini & Retzlaff, 1988). This matter should be kept in mind when discussing
the factor structure of the test.

Millon (1983a) initially reported a four-factor structure for the MCMI–I consist-
ing of general Maladjustment, Paranoid Behavior and Thinking, Schizoid Behavior
and Detachment, and Social Restraint/Aggression. For the MCMI–II, Millon (1987)
found an eight-factor structure but did not label them. Choca and Van Denburg
(1997) labeled these factors General Maladjustment, Acting Out/Self-Indulgent,
Anxious and Depressed Somatization, Compulsively Defended/Delusional–Para-
noid, Submissive/Aggressive–Sadistic, Addictive Disorders, Psychoticism, and
Self and Other Conflictual/Erratic Emotionality.

Subsequent MCMI factor studies with large sample sizes and a variety of clini-
cal populations (e.g., psychiatric inpatients and outpatients, substance abusers, and
criminal offenders) have generally employed a principal components analysis
with varimax rotation. Thus method variance is not a confounding issue in comparing results. Also, some of the studies used oblique rotations or a different factor method with no change in findings. Researchers have used different cutoff scores to identify the factor loadings. Some have used .30, but most have used .50 or even .60.

Despite these differences, there is a remarkable similarity of factor solutions reported across the studies. Usually from three to six factors have been reported. Most have found dimensions that can be called General Maladjustment, Schizoid Detachment versus Emotional Lability, Passive/Submissive versus Aggression, and Paranoid/Delusional Thinking.

To circumvent the problem of item overlap, researchers have also factored the items by taking the 175 items on the test and separating the 100 personality items from the 75 clinical symptom items. A factor analysis of these as separate entities should ascertain the latent variables that underlie the items themselves without the problem of items keyed on multiple scales. When this was done for the MCMI–I, results indicated six factors for the personality disorder scales and five factors for the clinical syndrome scales (Lorr, Retzlaff, & Tarr, 1989). When a similar methodology was used with the MCMI–II, researchers found six or seven factors associated with the personality disorder scales and anywhere from five to eight factors associated with the clinical syndrome scales, depending on the samples used (Lorr, Strack, Campbell, & Lamnin, 1990; Retzlaff, Lorr, Hyer, & Ofman, 1991).

Factor-based scales for the MCMI, using a variation of domain theory test construction, have also been developed (Retzlaff, 1993; Retzlaff & Gibertini, 1990). Eight narrow personality scales, five simplex clinical scales, three factor-based personality scales, and five broad psychopathology scales were constructed. They had high reliability across several populations and good validity estimates against the MCMI itself. Retzlaff (1993) argued that these scales provide a clear understanding of the common elements of the test.

A different approach to factoring is to conduct an analysis of each scale rather than all items on the test. When this was done for the MCMI–II, anywhere from one to five factors were found for each of the 13 scales as follows: Schizoid (Introversion), Avoidant (Aimless Emptiness and Introversion), Dependent (Submissiveness, Agreeableness, and Respectful Sociability), Histrionic (Behavioral Acting Out, Extroversion, and Agreeableness), Narcissistic (Self-Centered Arrogance, Extroversion, and Independent Superiority), Antisocial (Behavioral Acting Out, Sociopathic Independence, and Social Distrust), Aggressive–Sadistic (Emotional Acting Out, Strong-Willed Determination, and Social Independence), Compulsive (Affective Stability, Closeness to Experience, and Conscientiousness), Passive–Aggressive (Negativistic, Neuroticism, and Disagreeableness), Self-Defeating (Dysthymia, Sadomasochism, and Agreeableness), Schizotypal (Paranoia, Introversion, and Aimless Lassitude), Borderline (Depression, Behavioral Acting Out, Submissive Dependency, and Hostile Dominance), and Paranoid
Millon did not provide the factor structure of the MCMI–III (Millon, 1994, 1997a). The only factor study to date of this instrument was reported by Craig and Bivens (1998). They found a three-factor solution among 444 outpatients labeled General Maladjustment, Paranoid Behavior/Thinking with Detached Emotional- ity, and Antisocial Acting Out. These dimensions are similar to those previously found for the MCMI–I and MCMI–II when all the items are factored.

Aside from the basic science aspect provided by these results, they also have relevance to some aspects of Millon’s theory. Millon did not predict a basic factor structure that would correspond to the MCMI’s major divisions because normal, pathological, and symptom configurations are assumed to be interrelated in the- ory. Thus, the Schizoid, Avoidant, Schizotypal, and Dysthymic scales should load on a single factor, not only because of the item overlap in these scales but also because of the conceptual similarity of their underlying dimensions. The data in fac- tor analytic studies generally bear this out, although there have been some exceptions (O’Connor & Dyce, 1998).

Millon developed a domain-oriented schema for describing personality disor- ders (Millon & Davis, 1997). When individual scales are factored, the resulting solu- tions closely resembled Millon’s prototype description of each personality disorder in the domains of behavioral description, interpersonal conduct, cognitive style, expressive mood, and self-image. The areas of unconscious mechanisms, in- ternalized content, and intrapsychic organization were not well represented, al- though it would be hard to imagine how a self-report item could tap these elements (Choca, Retzlaff, Mouton, Strack, & Van Denburg, 1996).

The results of MCMI-based factor studies also pertain to classification issues in official psychiatric nomenclature. The DSM–IV (APA, 1994) classifies the 11 per- sonality disorders into three clusters. Cluster A is called the Odd or Eccentric dis- orders, Cluster B is called the Dramatic cluster, and the third group is named the Fearful cluster. MCMI factor data do not support these groupings (Bagby, Joffe, Parker, & Schuller, 1993), although some support for this classification scheme has appeared in the literature (O’Connor & Dyce, 1998).

In conclusion, although the underlying factor structure may be of theoretical in- terest, it is of little value clinically. We need content scales for each of the personality disorder and clinical syndrome scales that would allow us to determine the differential item endorsements in specific domains. This would greatly enhance our interpretations of the test.

Scale Reliability

A summary of empirical studies on the stability and reliability of MCMI scales has recently appeared in the literature (Craig, 1993a, 1997a). For the MCMI–I
there are 13 data sets for the Avoidant and Dependent scales and 12 data sets for the remaining scales. The test–retest interval has ranged from 5 days to 3 years, with most of the studies reporting a retest interval of about 3 months. For the MCMI–II there are eight data sets for the Dependent scale and seven for the remaining scales. The retest interval has ranged from 21 days to 4 months, with most of the studies retesting at about 2 to 3 months. For the MCMI–III there are two or three data sets, depending on the scale. The retest interval has ranged from 5 days to 6 months. The median correlations for clinical samples are reported later.

For the MCMI–I personality disorder scales, median correlations ranged from .19 (Passive–Aggressive) to .91 (Histrionic) with a median of .71 for all scales. For MCMI–I clinical syndrome scales, median correlations ranged from .45 (Somatoform) to .67 (Bipolar: Manic) with a median value of .60 for all scales. Median correlations for the MCMI–II personality disorder scales ranged from .62 (Borderline) to .78 (Compulsive), with a median of .74 for all scales. For the MCMI–II clinical syndrome scales, median correlations ranged from .43 (Somatoform) to .72 (Drug) with a median value of .66 for all scales. Since the initial summary reports (Craig, 1993a, 1997a), some additional data have become available (Craig & Olson, 1998; Piersma & Boes, 1997) on the stability of the MCMI–III scale scores as follows: For the personality disorder scales, median correlations (based on three data sets) ranged from .58 (Depressive) to .93 (Depressive) with a median of .78; for the clinical syndrome scales, median correlations ranged from .44 (PTSD) to .95 (Major Depression) with a median of .80.

The data suggest some interesting trends. Millon argued that the stability of personality disorder scales should be higher than the clinical symptom scales because personality is ingrained and more difficult to change and thus not susceptible to daily fluctuations. Clinical symptoms can wax and wane and also would be the focus of interventions designed to reduce those symptoms. These factors should produce lower reliability estimates in the clinical symptom scales and the data tend to bear this out across all versions of the test.

In general, the scale reliabilities have increased in stability with each edition of the test, suggesting more psychometrically sound scales with each revision. However subsequent studies have uniformly found lower reliability estimates among diverse populations compared to those reported in the test manuals. For the MCMI–I and MCMI–II, most of the scales seem sufficiently reliable over reasonable test–retest intervals with the possible exceptions of the Dependent, Passive–Aggressive, Borderline, and Paranoid scales. There is only a limited amount of research on the stability of MCMI–III scales, but preliminary evidence suggests improvement in stability estimates across most scales. These data are important because validity estimates cannot be higher than a scale’s reliability.
Detection of Profile Invalidity

Millon (1987) argued that

Deliberate misrepresentation ... on personality or clinical inventories is much less frequent than is commonly thought. Similarly, the role of “response styles” as a source of distortion seems to be a minor factor when compared with the content of substantive scales. (p. 195)

In spite of this belief, Millon took the approach of making computer adjustments on all MOCI scale scores that he believed to be affected by underreporting or overreporting of symptoms and problems. In a typical clinical setting, the vast majority of MOCI protocols require some form of correction. Scales Y (Desirability) and Z (Debasement) are used to detect and correct fake-good and fake-bad response sets. Test results are considered invalid when respondents endorse two or three of the unusual items on the three-item Validity Index, or obtain scores above or below a certain level on the Disclosure Index (Scale X). Scale X is a composite score derived from the raw scores of the 10 basic personality disorder scales. There are 14 studies that have investigated the MOCI validity scales, called modifier indices, to assess the ability of the MOCI to detect invalid profiles.

Results are only suggestive and have not been replicated with the MOCI-III. Most studies report that the MOCI has been able to detect fake-bad response sets with better efficiency than a fake-good response set (Bagby, Gillis, & Dickens, 1990; Bagby, Gillis, Toner, & Goldberg, 1991; McNeil & Meyer, 1990; Retzlaff, Sheehan, & Fiel, 1991; Van Gorp & Meyer, 1986). One study found that MOCI-II Disclosure and Debasement scales were positively correlated with the Sixteen Personality Factor questionnaire (16PF; Grossman & Craig, 1995) Faked-Bad scale and negatively correlated with the 16PF Faked-Good scale; the Desirability scale was positively correlated with the 16PF Faked-Good scale. In general, however, these preliminary findings are in line with research that has found that objective self-report inventories are more able to detect malingering and exaggeration than faked-good response sets (Baer, Wetter, & Berry, 1992; Berry, Baer, & Harris, 1991).

Diagnostic Accuracy

Because Millon raised the standard of psychological measurement by using precise scores for determining whether or not a patient has a certain disorder, there has been extensive research on the diagnostic accuracy of this instrument in selective categories. I highlight major trends in this research and the interested reader is
referred to other sources for more detailed presentation of these data (e.g., Craig, 1993a, 1997a; Millon, 1997b).

**Depression.** The assessment of affective disorders has been the subject of many MCMI studies. The test has three scales (Dysthymia, Major Depression, and Bipolar: Manic) and clear-cut decision rules to establish an affective disorder diagnosis. This specificity has spurred research on this topic. In particular, Wetzler and his colleagues have studied various depressive types with large samples of depressed patients. Results of these studies suggest few MCMI differences between unipolar and bipolar depressed inpatients. The unipolars had significant elevations on Dysthymia, Anxiety, Dependent, and Borderline scales on the MCMI–I, whereas the bipolars had elevations on Dysthymia, Anxiety, and Passive–Aggressive scales. For the MCMI–II, the unipolars had elevations on Dysthymia, Anxiety, Avoidant, and Dependent scales, whereas the bipolars had elevations on Dysthymia, Anxiety, Dependent, and Compulsive scales (Wetzler, Khadivi, & Oppenheim, 1995). Patients with Major Depression had significant elevations on Anxiety, Dysthymia, Somatoform, Passive–Aggressive, Borderline, Dependent, and Avoidant MCMI–I scales (Wetzler, Kahn, Strauman, & Dubro, 1989; Wetzler & Marlowe, 1993). Manic patients obtained elevated scores on Paranoid, Narcissistic, Hypomania, Drug Abuse, and Psychotic Delusion scales for MCMI–I and on Paranoid, Narcissistic, and Psychotic Delusion scales for MCMI–II. Neither Major Depression nor Hypomania was sufficiently sensitive to identify depressed and manic patients (Wetzler & Marlowe, 1993).

Although MCMI scales may be related to depressive disorders (Choca, Bresolin, Okonek, & Ostrow, 1988), most studies have found poor concurrent validity of the Major Depression scale for both the MCMI–I and MCMI–II tests. This was because there were no vegetative signs of depression in the scale, although that is the essential feature defining the disorder. Another problem has been the poor concurrent validity of the MCMI in detecting psychoses. To the extent that a patient has major depression with psychotic features, this would add to the diagnostic difficulty of the scale. These problems may have been corrected in the MCMI–III. The Major Depression scale now includes vegetative signs and we should see improved diagnostic accuracy with affective disorders.

**Substance abuse.** The MCMI has been used extensively to assess substance abusers, and this literature has been reviewed elsewhere (Craig & Weinberg, 1992a, 1992b). Research suggests that the modal profile among heroin and cocaine abusers is a 6A5 codetype, whereas variations of the 238A codetype are modal for alcoholics. However, cluster analysis reveals several subtypes. For drug addicts these subtypes include an antisocial/narcissistic type (the “parent” codetype), a
passive-aggressive/withdrawn style, a dependent type, and also a normal limit profile type (e.g., no personality disorder). For alcoholics the subtypes include an antisocial/narcissistic type, a negativistic/borderline type, a dependent/avoidant type, and a normal limit profile. A recent MCMI–III study with drug addicts has reaffirmed the 6A type as primary in this population (Craig & Olson, 1997). Both Scale T (Drug Dependence) and Scale B (Alcohol Dependence) show good stability over short intervals. Whereas Scale B shows good predictive power, no study has reported Scale T sensitivity levels above 50%, and there is evidence that some drug addicts can avoid the detection of their substance abuse on the MCMI (Craig, Kuncel, & Olson, 1994). However the sensitivity levels of both Scales T and B have improved compared to previous test versions (Craig, 1997b). A compulsive drinking style (e.g., elevated Compulsive) is associated with higher social functioning, episodic drinking, and higher abstinence rates among alcoholics, whereas elevations on Thought Disorder are associated with more debilitative forms of drinking. Use of the MCMI to assess substance abusers has recently been presented and the interested reader is referred to those sources for greater detail and specificity (Flynn, McCann, & Fairbank, 1995; Flynn & McMahon, 1997).

Posttraumatic stress disorder (PTSD). As with substance abuse, the MCMI has been frequently used to assess PTSD, primarily in Vietnam combat veterans. This literature shows that the modal profile of PTSD combat veterans is a codetype characterized by clinical elevations on Passive–Aggressive–Negativistic and Avoidant. This codetype has consistently appeared in research studies (Hyer, Melton, & Gratton, 1993), although some variations of this “parent” codetype have appeared in cluster analysis research (Hyer, Davis, Albrecht, Boudewyns, & Woods, 1994). A recent study using the MCMI–III among substance abusers with and without PTSD showed that the PTSD group had a modal profile characterized by elevations on Schizoid and Antisocial, which is a codetype substantially different than the PTSD codetypes using the MCMI–I and MCMI–II (Craig & Olson, 1997).

The research just presented pertains to men with PTSD. There is only one MCMI (MCMI–III) study among women with PTSD and the modal codetype had elevations in Depressive, Self-Defeating, and Avoidant scales (Allen, Coyne, & Huntoon, 1998). The MCMI–III has a scale that now specifically addresses PTSD and was designed for use in both military and civilian trauma. Craig and Olson (1997) found that this scale (R) was the best predictor of PTSD in a multiple regression equation and had sensitivity and specificity levels among drug addicts with and without PTSD at higher values than those reported in the test manual. We look forward to more PTSD research, particularly with women and civilian trauma patients, using the MCMI–III. The reader interested in more in-depth presentation on using the MCMI with PTSD is referred to a recent summary of this research (Hyer, Brandsma, & Boyd, 1997).
**Personality disorders.** The MCMI is primarily a diagnostic instrument for the assessment of personality disorders (Millon, 1996). Accordingly, the bulk of MCMI research has investigated its concurrent and convergent validity in assessing personality disorders. This literature is too vast to summarize. However, a few trends emerge. The MCMI shows moderate convergent validity with similar self-report inventories but low convergent validity with structured psychiatric rating scales. The Paranoid scale shows poor convergent validity with similar measures. There is no information on these matters using the MCMI–III.

Here I focus on three scales—Histrionic, Narcissistic, and Compulsive—that merit special attention. I argue that the bulk of the data suggest that Histrionic and Compulsive are measuring personality styles and not personality disorders, whereas Narcissistic may measure either a style or a disorder. The evidence is presented in what follows.

For the Histrionic scale, factor analytic research clearly shows that the scale loads positively with items pertaining to extroverted traits and behaviors and it loads negatively with items dealing with psychological maladjustment. Convergent validity studies show that this scale correlates positively with measures of mental health and correlates negatively with measures of emotional maladjustment. In fact some studies found that elevations on this scale are associated with less distress, more positive life events, and fewer social problems. When the MCMI is used with nonclinical samples, including Air Force pilots in training, female graduate students in psychology, and normal women, they often attain their highest score on this scale. Primary elevations on the Histrionic scale are infrequent among psychiatric samples. These data suggest the scale is measuring a histrionic personality style and not a disorder.

The exact same pattern is true for the Narcissistic personality disorder scale. Factor analysis shows it loads negatively with items dealing with maladjustment; it correlates with measures of mental health rather than with measures of maladjustment; many nonclinical samples (mostly men) obtain their highest score on this scale; and, with the exception of substance abusers, primary elevations on this scale are rare among psychiatric samples. However the scale also correlates positively with measures of pathological narcissism, particularly with the Narcissistic personality inventory and with the Narcissistic personality disorder on the Minnesota Multiphasic Personality Inventory (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Hence the clinical task is to determine whether elevations on Narcissism represent a style or a disorder.

The evidence is even stronger that the Compulsive personality disorder scale is actually measuring a compulsive style and not a disorder. The scale correlates positively with items suggesting control of behavior and emotions; correlates negatively with measures of psychiatric disturbance; is rare in the codetype of psychiatric patients; and appears as a frequent high point scale among nonclinical men, including Air Force pilot trainees, family practice residents, and male college
students. It shows low convergent validity with other measures of compulsivity, and elevated scores on this scale have been associated with better treatment outcomes and improved mental health. Finally, patients who had an obsessive–compulsive disorder did not have significant elevations on the Compulsive personality disorder scale. The evidence suggests that this scale is measuring a compulsive style but not a compulsive disorder.

Some have argued that the entire inventory is really measuring personality styles rather than disorders (Choca et al., 1992), but this view is not generally accepted.

Effects of Demographic Variables

The demographic variables most often discussed as possibly affecting test scores are race and gender. More recently the effects of culture are being considered as a moderating variable, but this has been inadequately studied in most psychological tests and there are no specific research studies that have addressed this issue with any version of the MCMI. The data summarized in the following sections are extremely tentative and based on up to six studies, all from the MCMI–I and MCMI–II. No information is presently available on the effects of moderator variables on MCMI–III scores.

Race. African Americans consistently score higher on Scales 5, 6A, P, T, and PP. Whites score higher on D. No racial differences appear between Blacks and Whites on Scales 3, 7, 8A, and A. No data are available on comparisons between Whites and other ethnic groups on MCMI scales.

Gender. Men score higher on Scale 6A; women score higher on Scales H and CC. No gender differences consistently appear on Scales 2A and 8A. No other conclusions are warranted from the data.

How is one to interpret the nature of these differences? One possibility of such patterns is test bias, but an equally plausible interpretation is that the test is tapping true differences in the populations. Furthermore, the diagnoses of the patients may not have changed, even when a group obtained statistically higher scores on a given scale. There simply are not enough data to answer the issue of test bias. As with any assessment instrument, one should weigh carefully the possibility of moderator variables as contributing factors in final behavioral expression.

STRENGTHS AND WEAKNESSES

The MCMI is one of the few assessment instruments in psychology that was derived from a comprehensive theory. It is increasingly coordinated with the multi-
axial format in the DSM. It enhances diagnostic efficiency by taking into account the base rates of the disorders it measures. It is also relatively quick to administer (20–30 min), and its scales are refreshingly easy to interpret once one learns the theory and has a grasp of DSM terminology. On the other hand, the test is susceptible to patients with an acquiescent response set because most of the items are keyed true. It does poorly in assessing patients with minor personality pathology and those with severe personality dysfunction (e.g., psychotic disorders). In my opinion, it should not be used as a broadband screening instrument because one must suspect in advance that there is pathology before the test is selected for use in a given clinical situation. Interpretation of personality scales in the BR range of 75 to 84 is a problem. At this level the person has some of the behaviors and traits that define a disorder but not all of them. However there is no way to determine which behaviors and traits are extant. There may be subtypes of the different personality disorders (e.g., a dependent borderline, a histrionic borderline, an antisocial borderline), but assessment of these subtypes has not been incorporated into the MCMI. Finally, there is a lack of a research base for the MCMI–III, which most assuredly will be corrected in coming years.

THE FUTURE

Future refinement of the MCMI is already in the planning stages (Millon & Davis, 1997). Under consideration are trait subscales for the personality disorders and a more explicit delineation of the kinds of symptoms and traits that can reasonably be expected to be validly assessed on a self-report instrument. Millon and Davis also argued for greater theoretical refinements in the disorders themselves, as instrumentation follows theory and typology. Finally, with the recent development of theoretically based treatment strategies (see Dorr, this issue), we can expect to see MCMI codetypes tied to specific interventions.

In conclusion, it is easier to criticize than to create. Millon is a brilliant theoretician who has created a number of clinically useful assessment inventories, perhaps the best of which is the MCMI. The three versions of the MCMI have generated a substantial amount of interest and research in the past 20 years, and many clinicians faithfully value its insights and findings. From the man, we can ask no more!

REFERENCES


Robert J. Craig
Illinois School of Professional Psychology
20 South Clark Street, 3rd Floor
Chicago, IL 60603

Received January 18, 1999